PERIODICAL ROOM
GENERAL LIBRARY
UNIV. OF MICH.

LB 1051 .J6 1931

# The Journal of EDUCATIONAL PSYCHOLOGY

*Devoted Primarily to the Scientific Study of Problems of Learning and Teaching.*

HAROLD RUGG, *Editor*

VOL. XXII, No. 2 FEBRUARY, 1931 $5.00 per year

## CONTENTS

PRACTICE VERSUS GRAMMAR IN THE LEARNING OF CORRECT ENGLISH USAGE. *Percival M. Symonds* . . . . . 81

A SCALE OF MILITARISM-PACIFISM. *D. D. Droba* . . . . . 96

WHAT THE THEORY OF FACTORS IS *Not*. *C. Spearman* . . . . . 112

A REPLY TO SOME RECENT CRITICISMS BY PROFESSOR SPEARMAN. *H. D. Carter* . . . . . 118

STUDIES IN HANDEDNESS: III. RELATION OF HANDEDNESS TO SPEECH. *R. H. Ojemann* . . . . . 120

THE CONDITIONING OF OVERT EMOTIONAL RESPONSES. *Harold Ellis Jones* . . . . . 127

CERTAIN AMBIGUOUS TERMS IN EDUCATIONAL PSYCHOLOGY. *Stephen Maxwell Corey* . . . . . 131

THE AUTOMATIC PREDICTION OF SCHOLASTIC SUCCESS BY USING THE MULTIPLE REGRESSION TECHNIQUE WITH ELECTRIC TABULATING AND ACCOUNTING MACHINES. *David Segel* . . 139

A TECHNIQUE FOR EXPERIMENTATION ON GUESSING IN OBJECTIVE TESTS. *Louis Granich* . . . . . 145

NEW PUBLICATIONS IN EDUCATIONAL PSYCHOLOGY AND RELATED FIELDS OF EDUCATION . . . . . 157

Published Monthly Except June to August by

**Warwick & York, Inc.,**

10 East Centre Street Baltimore, Md.

Entered as Second Class matter November 15, 1921, at the Post Office at Baltimore, Md., under the Act of March 3, 1879; additional entry as second class matter at York, Pa.

# THE JOURNAL OF EDUCATIONAL PSYCHOLOGY

ESTABLISHED 1910

*Editor:* HAROLD RUGG

*Board of Associate Editors*

J. CARLETON BELL — ARTHUR I. GATES — RUDOLF PINTNER — E. L. THORNDIKE
FRANK N. FREEMAN — V. A. C. HENMON — L. M. TERMAN — GOODWIN B. WATSON

*Assistant to the Editor:* FRANCES M. FOSTER

The Journal is published monthly, except June to August, by Warwick and York, Inc., Baltimore, Md., and York, Pa.

The price is $5.00 a year, payable in advance; foreign postage 30 cents extra. Single current issues are 70 cents each, and subscriptions for less than a year are charged at single issue rate for each number ordered. Back volumes are $6.00 each; single back numbers, 90 cents each, except, when more than three years old, the price is $1.00 each.

Subscribers should notify the publishers of change in address at least four weeks before publication of the issue with which the change is to take effect. Claims for non-receipt of an issue will not be honored unless made within two weeks after receipt of the next succeeding number.

Manuscripts for publication, books or other materials for review, and correspondence concerning editorial matters should be addressed to Harold Rugg, Editor, 425 West 123rd Street, New York City. Subscriptions should be sent to the publishers.

WARWICK AND YORK, INC., *Baltimore, Maryland*



**TEACHERS' HEADQUARTERS**

*European Plan*

*Entirely Fire-Proof*

*Centrally Located*

**EDWARD DAVIS**
*Manager*

**HOTEL RENNERT**
**BALTIMORE, MARYLAND**

The Psychology of Learning
1928 Revision — BY WILLIAM H. PYLE
$2.20 + 12¢ postage
Warwick & York, Inc., *Baltimore, Maryland*

***The Acid Test***

**What more exacting test can there be of dictionary leadership than the judgment of the educational world?**

**Every State that has adopted an unabridged dictionary for use in its schools has chosen exclusively**

**WEBSTER'S NEW INTERNATIONAL DICTIONARY**

**The Colleges voted overwhelmingly in favor of Webster as the standard of pronunciation, in answer to questions submitted by the Chicago Woman's Club.**

**Nearly 100% of all schoolbooks are based on the New International for pronunciation, spelling, compounding, and division of words.**

**The New International has been universally accepted by those best fitted to judge.**

**It is in accord with the best modern practice.**

**Use it in your own schools.**

***Write for our free booklets of interesting ways to teach the use of the dictionary.***

**GET THE BEST**

**G. & C. MERRIAM COMPANY**
**SPRINGFIELD, MASS.**

THE JOURNAL OF
EDUCATIONAL PSYCHOLOGY

Volume XXII | February, 1931 | Number 2

# PRACTICE VERSUS GRAMMAR IN THE LEARNING OF CORRECT ENGLISH USAGE[1]

PERCIVAL M. SYMONDS

Teachers College, Columbia University

Recent correlation studies have shown little relationship between knowledge of English grammar on the one hand, and the correctness of usage on the other. Those who are familiar with these studies commonly believe that the value of English grammar in influencing correct grammatical usage in speaking or writing has been definitely discredited. But correlation studies have frequently misled in the past regarding the actual dynamic relationships in learning and the amount of transfer that can be expected. Grammar still has a somewhat thriving existence in our schools.

In 1906 Hoyt[2] reported a correlation as low as .18 between grammar and composition ability among high school pupils. Rapeer[3] reports the same relationship to be .23. Segal and Barr[4] in a minor study reported a correlation of .48 between a test of formal grammar and a test of "applied grammar" which is really a test of ability to choose the correct form. Asker[5] reached similar conclusions by a different

[1] The author is indebted to Mr. Frederick Graham, formerly principal of P. S. 210 Brooklyn, Mr. Jacob Theobold, principal of P. S. 165 Manhattan, Mr. Abraham Ehrenfeld, principal of P. S. 10 Manhattan, and Mrs. Ruth D. Schatteles, formerly principal of P. S. 87 Manhattan, for their most helpful cooperation in this experiment, and to the many teachers who participated.

[2] Hoyt, F. S.: "The Place of Grammar in the Elementary Curriculum." *Teachers College Record*, Vol. VII, November, 1906, pp. 467–500.

[3] Rapeer, L. W.: "The Problem of Formal Grammar in Elementary Education." *Journal of Educational Psychology*, Vol. IV, March, 1913, pp. 125–137.

[4] Sega, D., and Barr, N. R.: "Relation of Achievement in Formal Grammar to Achievement in Applied Grammar." *Journal of Educational Research*, Vol. XIV, December, 1926, pp. 401–402.

[5] Asker, W.: "Does Knowledge of Formal Grammar Function?" *School and Society*, Vol. XVII, January 27, 1923, pp. 109–111.

method. The general conclusion reached from these correlation studies, to quote Segal and Barr, is that "formal grammar has no immediate transfer value so far as applied English grammar is concerned."

However, the true test is the experimental test. What real influence does learning English grammar have on usage? To determine what influence exists, the writer carried out some test-teach-test experiments in Grade VI of four New York City elementary schools.

The test which was used in the experiment consisted of forty items and was modeled after the Charters Diagnostic Language Tests. Each item (except the first) consisted of a sentence having a grammatical error to be corrected by the pupil by rewriting the sentence in the space beneath the item. Samples of the items are:

14. Are you most ready to go?

.....

28. He said that I was most there.

.....

Previous experience with the Charters test indicated that the lowest initial scores, and consequently those items of greater difficulty, were made on adjectives, adverbs, prepositions, and conjunctions. Consequently in planning for this experiment all of the items selected for this test were of these types. This test was called Test I, and was always given first as well as last in the experiment, with the hope of discovering what changes took place on the test due to the activities carried on in between.

The technique of the Charters test was retained as the best simple measurable evidence of usage. It would have been possible to have made the test a truer measure of choice in an uncontrolled situation, but probably at sacrifice of scorability and reliability. On the other hand in using a recognition test, pupils would have been more keenly attentive to differences in form and it would have too narrowly a test situation and not enough an opportunity to recognize spontaneously when a form is incorrect and what the correct usage should be.

In every case Test I was immediately repeated at the opening of the experiment, and all gains were computed from the second application of the test instead of the first in order to eliminate the influence of practice on the test. The average practice effect was found to be 1.14.

Six different experimental procedures were tried. Each procedure was conducted in three classes in a different school under very carefully controlled conditions. It was believed that in concentrating a pro-

cedure within a school there was less danger of having teachers discuss what they were doing and hence vary their procedure from that which was defined for them. No attempt was made to equate groups except the rough one of choosing Grade VI groups. This defect in experimental procedure is recognized. The possible influence which differences in selection in the groups might have on the results and attempts to allow for possible differences will be discussed. It would have been possible to carry out the study on the scale contemplated if the orthodox equating techniques had been followed.

### Experimental Procedure I: Repetition of Correct Forms

Four classes in each of two schools, P. S. 10 Manhattan and P. S. 210, carried out the experiment in pure repetition in 1929. Sheets were prepared which contained forty sentences each of which were grammatically correct. Each sentence contained a usage which which corresponded to a similar incorrect usage in Test I. Examples of items in this sheet of correct sentences (Exercise A) are as follows:

14. The boy was almost killed by an automobile.
28. My baby brother is almost two years old.

Exercise A was distributed to each child and directions to the teacher stated:

> Have the class read through the sentences aloud and in unison. Let the teacher first read the number of the item and then the class read the sentence. Do not let the children hurry. See that they read slowly and distinctly. See that everyone reads. Do not allow the class to repeat any of the sentences unnecessarily. If the class reads a sentence incorrectly or mispronounces a word, have them repeat it (once) correctly.

The sheets were immediately collected when this reading was finished. This device made possible one repetition of each correct form by each child.

In order to check up on the influences of different amounts of practice on this new test (Test I) as compared with one previous experiment, *one repetition* of Exercise A was given between the two applications of Test I in two classes; *three* repetitions between tests in two classes; *five repetitions* in two classes; and *ten repetitions* in two classes.

This part of the experiment was designed to give a measure of the influence of sheer repetition with which the results of other experimental procedures could be compared.

### Experimental Procedure II: Repetition of Correct and Incorrect Forms

In this section of the experiment carried out in P. S. 165 Manhattan in 1929, mere repetition was again studied, but in this case both the correct and incorrect forms were repeated side by side. This new exercise (Exercise A, Form 2) was distributed to the children with directions that it be read aloud and in unison as a class as in the previous procedure. Three repetitions of this exercise were interpolated between the tests and the results were comparable to the three repetitions used in the first experimental procedure.

Examples from Exercise A, Form 2:

| | |
|---|---|
| 14. The boy was most killed by an automobile<br>is wrong | 14. The boy was almost killed by an automobile<br>is right. |
| 28. My baby brother is most two years old<br>is wrong. | 28. My baby brother is almost two years old<br>is right. |

The pupils read *is wrong* or *is right* after each sentence.

### Experimental Procedure III: Knowledge of Definitions, Rules, and Principles and Grammar

In this section, carried out in P. S. 87 Manhattan in 1929, an attempt was made to determine the influence on usage (Test I) of mere learning of grammatical rules and principles. A grammar study guide was prepared containing definitions, rules, and principles which applied specifically to the errors in Test I.

It was an original intention to include in this first edition of the grammar guide only bare definitions and rules with no illustrations. But this seemed such a formal procedure for children in a classroom even in an experiment and so far removed from school procedure that illustrations were included. This experiment does not tell us how much the illustrations in illuminating the rules influenced the results.

Teachers were cautioned as follows:

Do *not* do any of the following:

Do *not* ask pupils to make up examples of their own illustrating the rules.

Do *not* give the pupils additional examples of the rules.

Do *not* ask pupils to analyze the rules.

Do *not* give the pupils practice in applying any of the rules.

The three teachers cooperating in the experiment were instructed to drill their pupils in memorizing the definitions, rules and principles so far as time permitted. Fifteen or twenty minutes a day for two weeks or more were used.

A sample of the grammar guide, as it applies to the particular items already given from the test, is given below.

*Most* and *almost* are often confused. *Most* is an adjective and should modify a noun.

Example: *Most* boats carry life preservers.

*Almost* is an adverb and should modify a verb.

Example: John has *almost* finished his work.

A sheet of "Questions on Rules for Correct Usage of Adjectives, Adverbs, and Conjunctions" was also supplied to each child.

Illustrations of items are:

22. What part of speech is *most?*
23. What part of speech is *almost?*

The teachers were instructed to use these sheets with the pupils in drilling them on these definitions, etc.

Finally a "Test of Knowledge of Rules for Correct Use of Adjectives, Adverbs, Prepositions, and Conjunctions" was prepared. This was to be given to the class after the study of the grammar was finished, but *before* final Test I.

Samples of items on this test on grammar rules are:

22. Most is an (1) adjective, (2) adverb, (3) preposition, (4) conjunction. ( )
23. Almost is an (1) adjective, (2) adverb, (3) preposition, (4) conjunction. ( )

The teachers were given the express goal of preparing their pupils for this test on grammar rules, using every means at their command, direct or indirect. They were to aim at one hundred per cent on the grammar rules test if possible and the extent to which they reached this goal was a measure of their success in teaching the grammar rules. That they were remarkably successful in this is evidenced by the fact that out of a total possible score of 36 on this test, the three classes *averaged* 33.1, 24.2, and 30.1.

The purpose of this work was to discover what influence the learning of the rules had on Test I.

### Experimental Procedure IV: Grammar Analysis

The fourth experiment was carried out in one class in P. S. 87 Manhattan and two classes in P. S. 210 Brooklyn in 1930. In this

part of the experiment the value of ability in the grammatical analysis of sentences in improving English usage was studied. Experiment IV went one step further than Experiment III and gave pupils practice in analyzing the grammatical construction of certain words and phrases. A new grammar guide was made which included not only the definitions, principles and rules as before, but also certain exercises in which pupils were expected to *recognize* and *name* the construction. In Experiment IV teachers were instructed *not* to do any of the following:

Do *not* have the pupils memorize or learn the rules or principles.
Do *not* give examples of or emphasize the *wrong* form in any of the exercises.
Do *not* give the pupils practice in applying any of the rules.
Do *not* have the children make up illustrations of their own.
Everything should concentrate on having children learn to analyze the sentences given.

Besides the definitions and examples in the grammar manual which are similar to those used in the previous experiment exercises were included of which the following are typical:

| | |
|---|---|
| 1. *Most* boats carry life preservers. | *Most* is an_____ and modifies_____. |
| 2. John has *almost* finished his work. | *Almost* is an_____ and modifies_____. |

A test of thirty-four items with a total possible score of sixty-one on "Grammar Analysis" was prepared and the teachers cooperating in this experiment were instructed to do everything possible to prepare their pupils to do well on this test. Their success is evidenced by average scores of 37.9, 42.6, and 42.8.

## Experimental Procedure V: Choice of Correct Constructions

Experiment V, which was conducted in P. S. 10 Manhattan in 1930, tested the value of practice in choosing correct constructions on English usage as measured by Test I.

A new grammar guide was again prepared including not only the necessary definitions, principles, rules, and illustrations, but also exercises in which pupils were given practice in choosing correct constructions. Samples of these exercises on the same distinction *most* and *almost* are:

*Exercises on Use of Most and Almost.*—Insert *most* or *almost* in the following blank spaces:

1. Are you_____over your cold?
2. _____birds build nests.
3. Which is the_____fun?

In this experiment the teachers were instructed to limit their teaching strictly to giving the pupils practice in choosing the correct constructions. They were told *not* to have the pupils memorize or learn rules or principles, *not* to give examples or emphasize the wrong form in any of the exercises, *not* to analyze the constructions in any of the examples or exercises, and *not* to have the children make up illustrations of their own.

As before, a test was prepared on the choice of correct constructions and teachers were expected to help prepare their pupils directly for this test. Items in the test were of the same kind as the practice exercises, illustrations of which have already been given. The success of the teaching may be seen from average scores of three classes. The scores were 24.3, 24.4, and 26.7 out of a possible 33.

In this procedure, as in the other, Test I was given first and last, the aim being to see what effect the practice in choosing correct constructions has on Test I.

## Experimental Procedure VI: Whole Program

In 1930 the whole program comprised in the previous experiments was tried out in P. S. 165 Manhattan. After Test I, the grammar manual was distributed and pupils in three Grade VI classes learned the various definitions, principles, and rules, practiced analyzing sentences for the constructions of certain words, and practiced choosing correct constructions. When all this was done tests were taken on knowledge of grammar rules, ability to analyze certain constructions, and ability to choose correct forms. This was capped by three repetitions of the right-wrong forms of Exercise A, Form 2, described in Procedure II. Then Test I was repeated. The average scores on the tests for three classes were as follows:

TABLE I.—AVERAGE SCORES ON KNOWLEDGE, ANALYSIS AND CHOICE TESTS GIVEN TO THESE CLASSES IN PROCEDURE VI

| | Knowledge test | Analysis test | Choice test |
|---|---|---|---|
| Total possible score | 36 | 61 | 33 |
| Class VIA | 34.2 | 55.4 | 31.4 |
| Class $VIB_1$ | 30.0 | 45.8 | 27.6 |
| Class $VIB_3$ | 30.1 | 33.1 | 26.7 |

TABLE II.—SCORES MADE ON TEST I BEFORE AND AFTER AN EXPERIMENTAL PROCEDURE TOGETHER WITH GAINS BY CLASSES AND AVERAGE GAINS

| Experimental Procedure I practice | One repetition | | Three repetitions | | Five repetitions | | Ten repetitions | |
|---|---|---|---|---|---|---|---|---|
| | P. S. 10 $VIA_3$ | P. S. 210 $VIB_{4R}$ | P. S. 10 $VIA_1$ | P. S. 210 $VIA_1$ | P. S. 10 $VIB_4$ | P. S. 210 $VIB_1$ | P. S. 10 $VIB_2$ | P. S. 210 $VIA_3$ |
| Second test | 5.80 | 15.87 | 10.44 | 11.69 | 8.33 | 22.45 | 6.48 | 7.59 |
| First test | 5.40 | 15.55 | 9.22 | 12.24 | 7.04 | 20.75 | 5.70 | 7.71 |
| Gain | +.40 | +.32 | +1.22 | −.55 | +1.29 | +1.70 | +.78 | −.12 |
| Average | +.36 | | +.34 | | +1.50 | | +.33 | |

| Experimental Procedure II, practice on right and wrong, P. S. 165 | $VIB_1$ | $VIB_2$ | $VIB_3$ |
|---|---|---|---|
| Second test | 24.77 | 21.70 | 24.48 |
| First test | 15.00 | 12.26 | 13.36 |
| Gain | + 9.77 | + 9.44 | +11.12 |
| Average gain 10.11 | | | |

| Experimental Procedure III, learning rules, P. S. 187 | $VIB_1$ | $VIB_3$ | $VIIA_1$ |
|---|---|---|---|
| Second test | 25.44 | 13.37 | 23.80 |
| First test | 17.20 | 10.26 | 17.48 |
| Gain | 8.24 | 3.11 | 6.32 |
| $\sigma_{\text{mean gain}}$ | .82 | .73 | .75 |
| Average gain 5.89 | | | |

TABLE II.—*Continued.*

| Experimental Procedure IV, analysis, P. S. 87 and P. S. 210 | P. S. 87 VIB | P. S. 210 $VIB_{4R}$ | P. S. 210 $VIB_{5R}$ |
|---|---|---|---|
| Second test | 23.90 | 16.28 | 15.02 |
| First test | 19.83 | 11.00 | 12.23 |
| Gain | 4.07 | 5.28 | 2.79 |
| $\sigma_{\text{mean gain}}$ | .84 | .59 | .59 |
| Average gain 4.05 | | | |

| Experimental Procedure V, choice, P. S. 10 | $VIB_1$ | $VIB_3$ | $VIB_4$ |
|---|---|---|---|
| Second test | 24.15 | 17.71 | 18.40 |
| First test | 14.50 | 10.16 | 11.58 |
| Gain | 9.65 | 7.55 | 6.82 |
| $\sigma_{\text{mean gain}}$ | .55 | .63 | .98 |
| Average gain 8.01 | | | |

| Experimental Procedure VI, total program, P. S. 165 | $VIA_1$ | $VIB_1$ | $VIB_3$ |
|---|---|---|---|
| Second test | 29.40 | 28.72 | 26.20 |
| First test | 13.45 | 16.03 | 14.72 |
| Gain | 15.95 | 12.69 | 11.48 |
| $\sigma_{\text{mean gain}}$ | .76 | .71 | .86 |
| Average gain 13.37 | | | |

Before discussing the significance of the results consideration must be given to possible extraneous factors which may have influenced the results. The groups were not equated, as was previously mentioned, except by the rough method of selecting Grade VI classes (one class was a VIIA section). It may be seen by inspecting the average scores on the first test in the above table that there are wide differences in initial language usage. It is possible that there are also differences in general ability, and it is concensible that the brighter classes might gain so much more from the experimental procedure than duller classes by virtue of their brightness as to invalidate the apparent differences.

These possibilities were investigated by finding the correlations which exist between various factors.

TABLE III.—COEFFICIENTS OF CORRELATION BETWEEN MENTAL AGE, IQ, INITIAL SCORE ON TEST I AND GAIN ON TEST I

| | Initial score-gain | MA-initial score | IQ-initial score | MA-gain | IQ-gain |
|---|---|---|---|---|---|
| Experimental Procedure III (*rules*) (average three classes) | ..... | ..... | +.672 | ..... | +.084 |
| Experimental Procedure IV (*analysis*): | | | | | |
| (Average three classes) | ..... | ..... | +.330 | ..... | +.157 |
| (Average two classes) | ..... | +.361 | ..... | +.076 | |
| Experimental Procedure V (*choice*) (one class) | ..... | ..... | +.415 | ..... | +.152 |
| Experimental Procedure III (*rules*): | | | | | |
| (Average three classes) | −.135 | | | | |
| (*Analysis*) IV (average three classes) | −.271 | | | | |
| (*Choice*) V (average three classes) | −.440 | | | | |
| (*Total*) VI (average three classes) | −.334 | | | | |
| Average | −.295 | | | | |
| Correlation for total group | −.138 | | | | |

These correlations indicate that there is a marked negative correlation between initial score on Test I and gain on the test. In other words pupils, making the lowest initial scores on the test make the greatest gain. This is probably an artifact of the test. Pupils with the higher scores do not have room enough on the test to make as great gains as those making lower scores. That this factor had no appreciable effect on the average gains *by classes* is seen from the following table. If this factor did influence average gains by classes, then the

TABLE IV.—AVERAGE INITIAL SCORE AND AVERAGE GAIN FOR EACH OF THE EXPERIMENTAL PROCEDURES

| School | Experimental procedures | | Average initial score | Average gain |
|---|---|---|---|---|
| P. S. 10, P. S. 210 | Practice | One repetition | 10.48 | + .36 |
| | | Three repetitions | 10.73 | + .34 |
| | | Five repetitions | 13.89 | + 1.50 |
| | | Ten repetitions | 6.71 | + .33 |
| P. S. 165 | Practice: Right and wrong | | 13.54 | +10.11 |
| P. S. 87 | Rules | | 14.65 | + 5.89 |
| P. S. 87 and P. S. 210 | Analysis | | 14.35 | + 4.05 |
| P. S. 10 | Choice | | 12.08 | + 8.01 |
| P. S. 165 | Total program | | 14.73 | +13.37 |

differences are really greater than they appear, for in this table there is a tendency for those classes with the highest initial average score to make the most gain.

There is also a marked correlation between intelligence and initial score in Test I. The brighter pupils tend to exhibit better language usage as measured by Test I. The correlation of gain and intelligence is negligible although it tends to be positive. Were it not for the artifact of the negative correlation between initial score and gain it is probable that the correlation between intelligence and gain would be much larger. Semi-partial correlations were computed between gain and IQ, eliminating initial score from gain,[1] showing that the correlation between gain and IQ is lower than it would otherwise be due to the nature of the test.

TABLE V.—SEMI-PARTIAL CORRELATIONS BETWEEN GAIN AND IQ ELIMINATING INITIAL SCORE FROM GAIN

| School | Class | Correlation IQ-gain | Correlation IQ-gain with initial score eliminated from gain |
|---|---|---|---|
| P. S. 87 | $VIB_1$ | −.337 | +.051 |
| | $VIB_2$ | +.676 | +.533 |
| | $VIIA_1$ | −.086 | +.158 |
| P. S. 87 | VIB | +.379 | +.419 |
| P. S. 210 | $VIB_{4R}$ | −.046 | +.235 |
| | $VIB_{5R}$ | +.137 | +.163 |

The influence of general ability of a group may have had some influence on the gains, but probably not much. The correlation is positive but low. The classes in P. S. 165 had high IQ's on the average and in both cases made the highest gains. On the other hand both

[1] By a semi-partial correlation is meant the elimination of the effect of a third variable from one of the variables. In this case the formula used was

$$r_{(12)3} = \frac{r_{13} - r_{12}r_{23}}{\sqrt{1 - r_{12}}}$$

where

1 = gain
2 = Initial score
3 = IQ

Semi-partial coefficients of correlation were first developed by Franzen. See Franzen, R.: A Comment on Partial Correlation. *Journal of Educational Psychology*, Vol. XIX, 1928, pp. 194–197.

of the classes in P. S. 210 participating in Experimental Procedure IV (analysis) were rapid advancement classes with high IQ's but they made comparatively low gains. Again the classes in P. S. 10 participating in Experimental Procedure V (choice) had average IQ's which were really low in this experiment and yet their gains were high. It is not believed that the fact that classes differed in initial ability had appreciable influence on the results.

## Interpretation of the Results

1. The results show that mere repetition of correct forms cause small gains on the test. It is possible that the small gains of 0.3 for one, three and ten repetitions of the correct forms could be explained as mere practice effect. The original practice effect between the first and second trials on Test I was 1.14, and it is possible that there should be another small practice effect gain between the second and third trials. At all events, when compared with the other procedures mere repetition of correct forms is very inferior.

2. If one recalls the gains found for *motivation* as described in a previous experiment[1] it will be seen that they sink into greater insignificance than appeared in the earlier article. Motivating devices apart from the learning experiences offered are practically impotent.

3. It will be seen from Experimental Procedure IV that mere learning of definitions, rules and principles of grammar yielded an appreciable gain on the usage test. This result was wholly unexpected and is hard to explain. It was thought before beginning the experiment that mere memorization of grammar rules had little or no influence on usage. The experiment, however, indicates that somehow or other these Grade VI children extracted enough meaning from the formal and abstract definitions and rules with which they were previously unacquainted to make more correct changes on the test than they did the first time. The positive correlations of gains with intelligence would indicate that this "transfer" was one that the brighter children profited by more than the duller children. Within the limited use of language as found in this experiment it is possible to say that mere learning grammar rules *does* improve usage.

4. Training in the analysis of grammatical constructions and practice in the choice of correct constructions each taken alone caused improvement in English usage. The development of ability to analyze

[1] Symonds, P. M. and Chase, Doris Harter: Practice vs. Motivation. *Journal of Educational Psychology*, Vol. XX, 1929, pp. 19–35.

sentences for grammatical constructions apparently had about the same influence on usage that memorization of the rules had. Practice in choice of correct expressions had measurably greater influence than the learning of rules or analysis.

5. Mere repetition of the right and wrong forms in succession so that the distinction between the two is clearly brought out, and so that it is definitely stated which is the right form, had greater influence on usage than any of the work with grammar.

6. The whole program including memorization of rules, practice in analysis of grammatical constructions, choice of correct forms and mere repetition of correct and incorrect forms in succession yielded results in improved usage better than any one single method alone. It is estimated that these Grade VI classes made a gain in correct usage which is ordinarily to be expected in three years by working on this program fifteen minutes a day for three weeks. The experiment has not included subsequent testing to determine the permanence of these gains. Presumably there is the usual falling away and the re-adoption of habits prevalent in home and community.

## Discussion

One thing stands out clearly: It is the quality of drill or practice that counts and not its amount. Mere mechanical repetition apparently yields almost no learning. Instead of increasing the amount of drill in school subjects, more attention should be paid to the nature of the drill. The best results come by making what is to be learned easily identifiable as Thorndike has emphasized in recent lectures.

This study should not necessarily be interpreted as a justification of the teaching of English grammar. Grammar does have an influence on usage, but at what cost? Without doubt for most children the difficulty and trouble of learning grammar as a means for improving usage is so great that more direct attacks on usage are certainly more profitable. Our own results show that mere repetition of correct forms, where it is clearly indicated what is correct and what the critical point at issue is, has more influence on usage than any procedure with grammar.

For gifted children the study of grammar may be profitable. In the first place, the gifted child learns grammar relatively more easily. In the second place, with general principles to work from, a knowledge of grammar will cover more cases. Whoever really comprehends a generalization has command over a much wider range of

specific applications. But the average child finds it difficult to learn the definitions, rules, and principles, and finds it difficult to apply them when learned. For most children learning correct usage is more profitably accomplished when the specific items are learned directly. The child who has learned by direct practice to distinguish right from wrong usage is probably better equipped to learn the grammar which gives a logical explanation of the habits of correct usage which he has formed. In this sense grammar is a summary or epitome of the usage which has already been learned directly, rather than a tool for guiding the learning of correct usage. In the writer's opinion it is relatively unprofitable for the average child either to study grammar as a means for learning correct usage (because there are more direct, simpler, and easier means for learning usage) or as a means of summarizing the correct usage which one has learned (because there are other things more worthwhile in the curriculum).

In all the experimental work it is tacitly understood that pupils are making some effort to do well in school, to do well on tests, to please the teacher, to do what other children are doing, to learn new things, and the like. Unless there was some driving force the pupils would never have taken the tests, or attended to or learned the exercises. Our previous experiment showed that additional stimuli to learning cannot be imposed from without. But there must be a minimum of driving force which can be depended upon to direct the children's attention to what is to be learned. The selection of material in harmony with pupil's *interests* and *previous learnings* is certain to produce better learning results than when interests and previous learnings are disregarded.

It is always questionable as to how far the results of a single experiment can be extended to apply to dissimilar situations. The writer likes to believe that these experimental results apply to problems of conduct in general. This experiment demonstrates that mere learning rules or codes has a direct influence on conduct, more for the brighter children than for the duller children. Apparently these children extract enough from the formal rules to apply to their own behavior. The practical problem, however, is not Can rules function in conduct? but What is the most economical method of guiding or controlling conduct? This depends partly on the person's ability to comprehend generalizations and apply them to specific situations. Probably for the majority of people the simple direct method of telling them what is the "right" or best thing to do will produce the best results. And

if it is not known or not certain that one thing rather than another is right or best, or if it depends on the circumstances, then it is preferable either in the first instance to teach nothing, or in the second instance to give considerable help in solving the problem which will fit the response to the circumstances.

Finally, a word as to the experimental attack on problems of this type as compared with correlation methods. Correlation analysis is essentially static. It merely reports the results of whatever methods or combination of methods have produced the present situation. Correlation analysis does not begin to exhaust the possibilities in method—it does not even experiment with method. Correlation analysis is extremely limited insofar as its guidance for education is concerned. The point has been reached where the correlation method has about exhausted its possibilities. The time has come for a more earnest use of the experimental approach to problems of learning and education.

However, if the experimental method is to be used it must give heed to the law of the single variable. Every true experiment in education presents an artificial educational situation. Time and again during the course of the work reported in this article, principals and teachers would complain that the methods which were being used were too formal and dry to be practically used. To this I would agree. To most persons an educational experiment must make use of the best conceived organization of methods and materials possible. Not so. The actual educational situation is too complex to be used in experiment. When one has compared one actual classroom method with another, one does not know to what specific factor to attribute the results. In true experimental work a single factor must be pulled out of its setting and alone be allowed to vary. Only in this way is it possible to build up educative procedures that are based on correct principles.

# A SCALE OF MILITARISM-PACIFISM[1]

D. D. DROBA

University of Chicago

## The Problem

The purpose of this article is to describe the construction of a scale for the measurement of militarism-pacifism. The term "militarism-pacifism" is used here to specify a particular attitude. In a very broad sense it denotes a predisposition to act with reference to the issue of war *vs.* peace.

To measure militarism-pacifism directly is impossible. One cannot enter into an immediate contact with the predispositions of other persons. As an indirect measure of attitude the verbal expression of a person, in the form of statements, was chosen. The statements cover the following topics in this field: Causes of war, purposes of war, results of war and peace, what is to be done at present about war and peace, what is to be done in case of war, and general judgments about peace and war.

We can assume that there is some relation between action and verbal expression. That is, a person who endorsed militaristic statements would probably tend to act like a militarist, and a person who endorsed pacifistic statements would be likely to act like a pacifist. However, no assumption is made as to the extent of relation between the verbal expression and action. If it is discovered that a person has endorsed statements indicating a pacifistic attitude, one cannot be sure that he will also act exactly in accordance with his endorsements.

In the present study verbal expressions or opinions will be used to designate the successive steps on the militarism-pacifism scale. In constructing the scale it will be desirable to find some way of determining the distances between the successive steps on the scale. It will be desirable, furthermore, that the distances between the steps be approximately equal. How to equalize these distances constitutes the chief problem.

To summarize, the main problem of this experiment is to construct a scale of militarism-pacifism with approximately equal steps on the

---

[1] The scale, in a modified form, together with instructions for its use, was published by the University of Chicago Press.

scale—the steps to be represented by statements of opinion concerning war and peace.

## The Method

A set of two hundred thirty-seven statements expressing various degrees of militarism-pacifism was collected from several sources. These sources may be divided into three main divisions: The literature on war and peace, statements about the issue of war *vs.* peace written by one hundred undergraduates and twenty graduates, and a number of statements devised and modified by the writer. Out of the two hundred thirty-seven statements sixty-seven longest, least clear, and least relevant statements were discarded. Three professors and the writer then went over the remaining statements, and as a result forty more statements were eliminated, leaving one hundred thirty statements for use in the experiment.

Three hundred students attending the University of Chicago were used as subjects. Among them were eleven unclassified, eleven freshmen, ninety sophomores, fifty-seven juniors, thirty-nine seniors, forty-two year graduate students, thirty second year graduates, seventeen third year graduates, and three Ph.D.'s. Most of the subjects were students taking elementary psychology. Other departments which also assisted were sociology, political science, divinity school, and education. Only voluntary service was asked for except in a few cases when the instructor made the experiment a substitute for some class exercise.

The presentation of the material was made in some cases by the experimenter, in others by an instructor. Generally the envelopes containing the material were distributed first during the class hour. The experimenter then read the instructions while the subjects followed him in reading their own sheets. Any questions asked by the students at this time were answered. The subjects were then asked to do the task at home and to return the envelopes with the material complete in a few days.

When the student opened the envelope he found in it the following material: A sheet of instructions, one hundred thirty statements each on a separate slip 5½ inches by 4½ inches, eleven index slips of the same size, each labeled with a Roman number I–XI, and a slip of similar size for obtaining the necessary information about the subjects.

The subject was instructed to lay out before him the eleven slips. He then was asked to put on slip I those statements which he believed express the most extremely militaristic opinions. On slip XI he was requested to put those statements which he believed express the most extremely pacifistic opinions, and on slip VI to put those which he thought express neutral opinions. On the rest of the slips he was instructed to arrange the statements in accordance with the degree of militarism and pacifism expressed in them.

Only about seventy-five per cent of the envelopes were returned. This automatically eliminated some of the most careless individuals. A few envelopes were returned unused which was preferable to a thoughtless and hurried performance. The index slips and the instruction sheets were returned in most cases and were used over again with the newly prepared sets of statements.

The frequencies of judgments obtained from the three hundred students for each statement were tabulated. Table I shows the frequencies of judgment only for statements chosen to constitute the final scale. These frequencies were then cumulatively added and the percentages of the total three hundred judgments determined for each of the resulting sums. Three samples are shown in Table II. It is evident from the second column that statement number 9 was put in group I (first column) by two hundred fifty-four subjects. Two hundred fifty-four is 84.6 per cent of the three hundred subjects. Twenty-eight people put the statement in group II. Adding twenty-eight to two hundred fifty-four, we get two hundred eighty-two which is ninety-four per cent of three hundred. Seven people put the statement in group III. Seven and two hundred eighty-two is two hundred eighty-nine; that is, 96.3 per cent of the total number of subjects or judgments. Percentages thus obtained were called cumulative proportions.

The same method of calculation was used for all the one hundred thirty statements. In order to show how this method applies to all kinds of distributions the three statements, 9, 10, and 8, were purposely selected to represent three completely different distributions. Statement 9 was put by most people to the extreme left, statement 10 in the middle groups, and statement 8 in groups lying to the extreme right (see Table II).

From data such as found in Table II the cumulative proportions were plotted against the eleven groups or degrees of militarism-pacifism. Sample curves are shown in Fig. 1. Militarism-pacifism

TABLE I

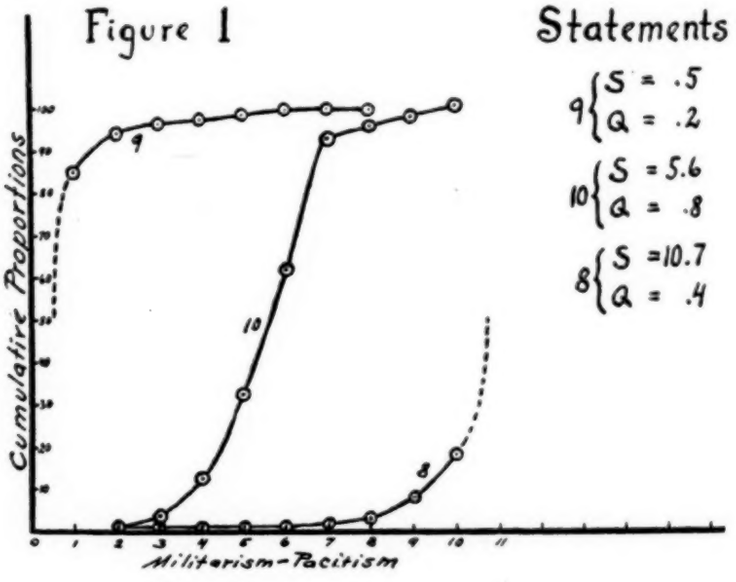
| Statements | Groups | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | I | II | III | IV | V | VI | VII | VIII | IX | X | XI |
| 1 | 11 | 56 | 81 | 86 | 51 | 10 | 1 | .. | 1 | 2 | 1 |
| 2 | ... | .. | 2 | 1 | 1 | 2 | 15 | 32 | 56 | 92 | 99 |
| 3 | ... | 3 | 25 | 31 | 115 | 97 | 22 | 5 | 1 | 1 | |
| 4 | ... | 2 | 2 | 5 | 2 | 8 | 46 | 80 | 85 | 41 | 29 |
| 5 | 1 | 3 | 1 | 3 | 1 | 10 | 18 | 36 | 76 | 92 | 59 |
| 6 | ... | 1 | 13 | 40 | 106 | 27 | 56 | 31 | 14 | 10 | 2 |
| 7 | 145 | 93 | 35 | 20 | 5 | ... | ... | 1 | .. | 1 | |
| 8 | ... | 1 | .. | .. | ... | ... | 2 | 5 | 15 | 31 | 246 |
| 9 | 254 | 28 | 7 | 3 | 4 | 3 | | | | | |
| 10 | ... | 2 | 8 | 26 | 60 | 88 | 93 | 9 | 6 | 8 | |
| 11 | 2 | 3 | 7 | 13 | 28 | 20 | 180 | 33 | 9 | 5 | |
| 12 | 13 | 36 | 47 | 86 | 80 | 33 | 1 | .. | .. | 3 | 1 |
| 13 | 6 | 18 | 38 | 69 | 105 | 54 | 4 | 2 | 1 | 3 | |
| 14 | ... | 2 | 4 | 5 | 16 | 123 | 55 | 34 | 32 | 20 | 9 |
| 15 | 5 | 1 | .. | .. | 1 | 1 | 4 | 8 | 13 | 55 | 212 |
| 16 | 2 | 1 | 3 | 6 | 16 | 45 | 78 | 70 | 40 | 24 | 15 |
| 17 | 102 | 87 | 58 | 36 | 13 | ... | 2 | .. | 2 | | |
| 18 | 1 | 1 | .. | .. | ... | 2 | 14 | 27 | 38 | 70 | 147 |
| 19 | 73 | 74 | 77 | 46 | 26 | 3 | ... | 1 | | | |
| 20 | 34 | 69 | 79 | 65 | 42 | 8 | 1 | 1 | .. | .. | 1 |
| 21 | ... | 1 | 2 | 2 | ... | 10 | 38 | 55 | 96 | 72 | 24 |
| 22 | ... | .. | .. | .. | 3 | 22 | 78 | 74 | 72 | 40 | 11 |
| 23 | ... | .. | 2 | 1 | 1 | 1 | 22 | 49 | 84 | 84 | 56 |
| 24 | ... | 7 | 21 | 69 | 131 | 55 | 11 | 2 | 1 | 3 | |
| 25 | ... | .. | .. | 1 | 3 | 46 | 65 | 76 | 55 | 43 | 11 |
| 26 | ... | .. | .. | .. | ... | 1 | 4 | 10 | 19 | 58 | 208 |
| 27 | 42 | 65 | 78 | 62 | 34 | 13 | ... | .. | 2 | 3 | 1 |
| 28 | 194 | 67 | 21 | 13 | 4 | 1 | | | | | |
| 29 | ... | .. | .. | 1 | 2 | 7 | 43 | 66 | 75 | 62 | 44 |
| 30 | 65 | 76 | 66 | 52 | 29 | 7 | ... | 1 | 2 | 1 | 1 |
| 31 | 33 | 50 | 72 | 68 | 59 | 13 | 1 | 2 | .. | 2 | |
| 32 | ... | 2 | 1 | .. | 5 | 29 | 53 | 73 | 74 | 49 | 14 |
| 33 | ... | .. | 1 | 1 | ... | 2 | 23 | 25 | 58 | 85 | 105 |
| 34 | 2 | 1 | 2 | 5 | 10 | 163 | 35 | 32 | 26 | 18 | 6 |
| 35 | ... | .. | .. | .. | ... | ... | 4 | 12 | 18 | 40 | 226 |
| 36 | 1 | 17 | 44 | 70 | 109 | 35 | 16 | 2 | 2 | 4 | |
| 37 | ... | 2 | 2 | 14 | 30 | 65 | 98 | 50 | 23 | 15 | 1 |
| 38 | ... | .. | 1 | .. | 2 | 4 | 21 | 18 | 33 | 65 | 156 |
| 39 | 93 | 85 | 62 | 36 | 20 | 1 | ... | .. | 1 | .. | 2 |
| 40 | ... | .. | .. | 2 | 4 | 29 | 42 | 67 | 78 | 62 | 16 |
| 41 | 16 | 40 | 64 | 82 | 76 | 14 | 3 | 2 | 1 | 2 | |
| 42 | 148 | 81 | 43 | 13 | 12 | ... | ... | .. | 2 | 1 | |
| 43 | ... | 1 | 3 | 1 | 3 | 67 | 89 | 54 | 46 | 30 | 6 |
| 44 | 6 | 8 | 19 | 22 | 35 | 208 | 2 | | | | |

TABLE II.—CALCULATION OF CUMULATIVE PROPORTIONS

| Number of the statement | 9 | | 10 | | 8 | |
|---|---|---|---|---|---|---|
| Groups | Cumulative frequency | Per cent | Cumulative frequency | Per cent | Cumulative frequency | Per cent |
| I | 254 | 84.6 | | | | |
| II | 28 | | 2 | | 1 | |
| | 282 | 94 | | .6 | | .3 |
| III | 7 | | 8 | | 0 | |
| | 289 | 96.3 | 10 | 3.3 | 1 | .3 |
| IV | 3 | | 26 | | 0 | |
| | 292 | 97.3 | 36 | 12 | 1 | .3 |
| V | 4 | | 60 | | 0 | |
| | 296 | 98.6 | 96 | 32 | 1 | .3 |
| VI | 3 | | 88 | | 0 | |
| | 299 | 99.6 | 184 | 61.3 | 1 | .3 |
| VII | 0 | | 93 | | 2 | |
| | 299 | 99.6 | 277 | 92.3 | 3 | 1 |
| VIII | 0 | | 9 | | 5 | |
| | 299 | 99.6 | 286 | 95.3 | 8 | 2.6 |
| IX | 1 | | 6 | | 15 | |
| | 300 | 100 | 292 | 97.3 | 23 | 7.6 |
| X | | | 8 | | 31 | |
| | | | 300 | 100 | 54 | 18 |
| XI | | | | | 246 | |
| | | | | | 300 | 100 |

was represented on the base line by the eleven groups running from extremely militaristic groups to extremely pacifistic ones. The cumulative proportions were located on the y-axis running from zero to one hundred.

Sample curves are shown in Fig. 1. Militarism-pacifism was represented on the base line by the eleven groups running from extremely militaristic groups to extremely pacifistic ones. The cumulative proportions were located on the y-axis running from zero to one hundred.

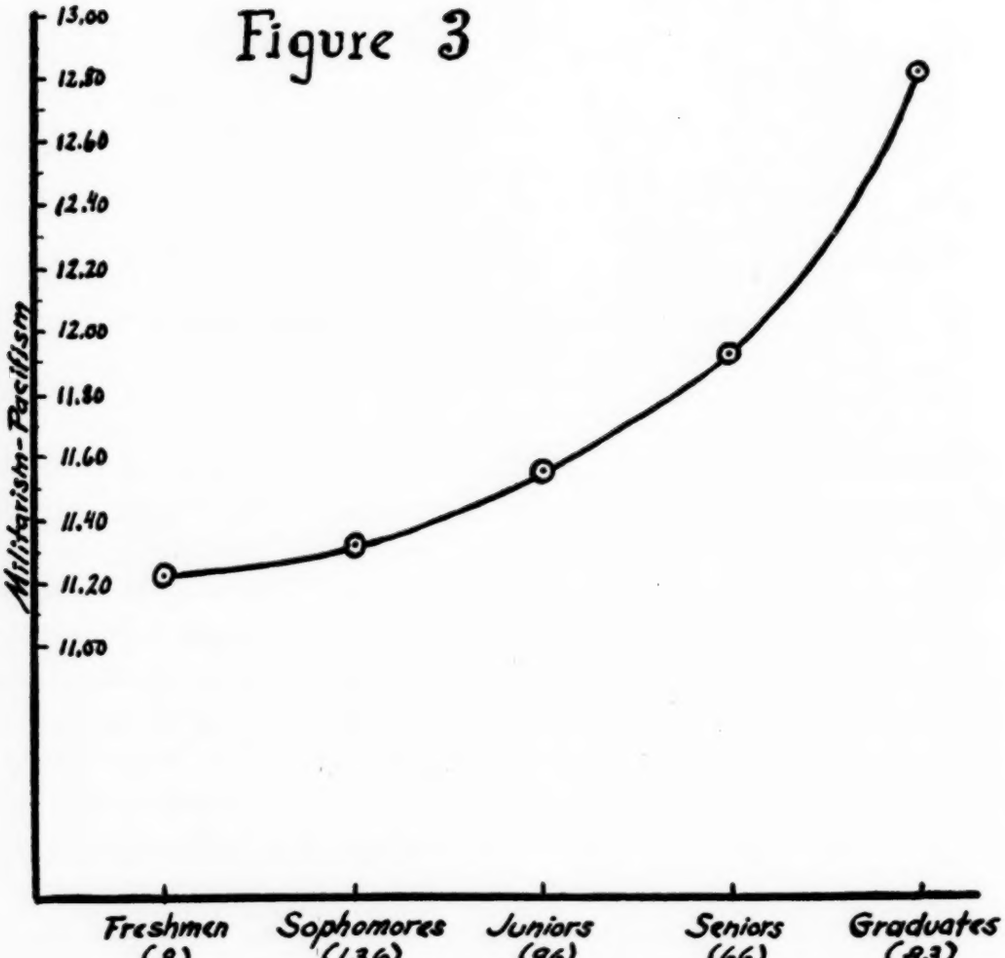



From the curves thus obtained two values were determined: The scale value and the quartile deviation. The scale value of a statement was found as follows: Somewhere on the chart the curve intersects the fifty per cent horizontal line, that is a line running horizontally at the level of fifty per cent and at right angles to the y-axis. From the point of intersection a line was dropped at right angles to the x-axis. The point where this line intersects the base line determined the scale value of the statement. The scale value of statement 10 (see Fig. 1) was found to be 5.6 and recorded on the right-hand side of the figure. A similar procedure was followed in finding the scale value of each of the one hundred thirty statements.

With some of the extreme statements, however, the curve did not reach the fifty per cent line. Examples of this may be found in Fig. 1 and Fig. 3. In such cases the curve was projected at sight down

Figure 2

Figure 3



or up as far as the fifty per cent line. Naturally, such procedure is not quite as accurate as when the curve reaches the fifty per cent line, but it is the best method available for the extreme curves.

Variability of a statement was measured by the semi-interquartile range or quartile deviation, designated by $Q$. The greater the value of $Q$, the more variable the statement, because people disagree more as to where the statement belongs. $Q$ value was obtained by simply reading off the range on the base line corresponding to the distance between the two points where the curve intersected the twenty-five per cent and the seventy-five per cent horizontal lines, and dividing this distance by two. The variability of statement 9 was found to be 0.2, the variability of statement 10 was 0.8, and that of statement 8 was 0.4 (see Fig. 1).

## The Scale

1. *Construction.*—The next task was the selection of statements to constitute the final scale. Two major criteria used for this purpose were (*a*) the scale values and (*b*) the $Q$ values. All of the one hundred thirty statements were arranged first on the basis of the scale values in ascending order, ranging between 0.5 and 10.7. Nearly every step in this range was represented by several statements of practically identical scale values. All statements of identical scale values were arranged according to their $Q$ values.

It was decided that there would be twenty-one steps in the entire scale, each step to be represented by one statement. An additional statement was to represent the origin. Approximately two statements were picked from each of the eleven groups. But in so doing the experimenter was guided by the scale values of each statement rather than by the groups.

In case of identical scale values the least variable statements were picked as measured by the $Q$ value. In a few instances both the scale values and the $Q$'s were identical. Two minor criteria were used in cases of this type: The brevity of a statement and the shape of the curve of its distribution. A regular curve was preferred to a less regular one and a short statement to a longer one.

As a result, a series of twenty-two statements were obtained that were about equally spaced along the scale. The distance between two adjacent statements seemed to the three hundred judges, on the whole, as large as the distance between any other two adjoining state-

ments. In terms of scale values none of these distances was larger than .6, and none smaller than .4 with the exception of the distance between the last two statements, which was .3. A second form of the scale was selected by examining the scale values and the *Q*'s again and picking equally spaced and least variable statements. Twenty-two statements were selected to match the first form as nearly as possible.

The scale was readjusted on the basis of the frequencies of endorsement obtained from four hundred students. The frequencies of endorsement were cumulatively added, and each sum in one of the forms was compared with the corresponding sum in the other form. If one of the sums was larger, the statements were exchanged so as to balance the frequencies in one form with the frequencies in the other.

After the two forms were thus equalized the scores were calculated for each and all of the four hundred subjects. The distributions of the frequencies of scores obtained from the two forms are shown in Fig. 2.

2. *Description.*—Forty-four selected statements constituted the main body of the scale in two forms. Thus each of twenty-one steps on the scale was represented by a pair of statements. One pair, the most extremely militaristic statements, represented the origin and were designated by *O*. The two most extremely pacifistic statements were designated by number twenty-one. In order to assure greater attention and careful reading on the part of the subjects, the statements were arranged in random order. Written instructions were found at the beginning of the scale. Attached to the scale was the information sought about those to whom the scale was given.

The instructions, the scale statements, and the information section are reproduced below. In front of each statement code numbers are found. After each statement in the parenthesis three numbers are given: The scale value, the quartile deviation, and the equivalent number (about double the scale value) of that statement. The latter number was used in scoring. In giving the scale these three numbers are, of course, omitted.

### A Scale of Militarism-Pacifism

#### Instructions

The purpose of this experiment is to measure differences in opinions relating to peace and war. Below you will find a number of statements expressing various degrees of opinion about militarism-pacifism.

If, on the whole, you agree with the sentiment of a statement, mark it with a *plus sign* in the parenthesis.

If, on the whole, you disagree with the sentiment of a statement, mark it with a *minus sign* in the parenthesis.

Please vote for each statement even if it seems ambiguous and you have to make some sort of guess.

Only mass results will be published which will not distinguish individual opinion.

Statements

( ) 1. Multitudes are benefited by learning the lesson of war-time discipline. (3.0, 0.9, 5)
( ) 2. War in the modern world is as needless as it is suicidal. (9.5, 0.9, 18)
( ) 3. Those military units should be retained which afford training to the body and mind. (4.8, 0.6, 8)
( ) 4. Army discipline injures self-respect and individuality. (8.0, 0.9, 15)
( ) 5. The losses in human life and property, great as they are, are small evils compared to the undermining of morals and the lowering of standards of culture and civilization caused by war. (9.0, 0.9, 17)
( ) 6. Wars are justifiable only when waged in defense of weaker nations. (5.2, 1.1, 9)
( ) 7. War is ennobling and stimulative of the highest and best qualities of humankind. (1.0, 0.8, 1)
( ) 8. It is the moral duty of the individual to refuse to participate in any way in any war, no matter what the cause. (10.7, 0.4, 21)
( ) 9. Might is right. (0.5, 0.2, 0)
( ) 10. The most that we can hope to accomplish is the partial elimination of war. (5.6, 0.8, 10)
( ) 11. The evils of war are slightly greater than its benefits. (6.5, 0.5, 12)
( ) 12. Until an equivalent discipline is organized war must have its own way. (3.6, 1.0, 6)
( ) 13. Because right may be more important than peace, war may be the lesser of two evils. (4.2, 0.9, 7)
( ) 14. Along with patriotism world-citizenship should be taught in all secondary schools. (6.0, 1.0, 11)
( ) 15. Persons should pledge themselves never to aid any future war. (10.4, 0.5, 20)
( ) 16. It is almost impossible to have a large military force without being tempted to use it. (7.0, 1.0, 13)
( ) 17. War is the tonic of races. (1.5, 1.1, 2)
( ) 18. When the next war breaks out we should tell the diplomats who lead us to it that we will not follow them. (10.0, 1.1, 19)
( ) 19. Militarism is necessary for the proper defense and protection of the individuals of a country. (2.0, 1.0, 3)
( ) 20. So long as any people, white, black, brown or yellow, hold weapons in their hands, we must not commit the folly of disarming. (2.5, 1.0, 4)
( ) 21. Under the scourge of war a nation has no opportunity for cultural development. (8.4, 0.9, 16)

( ) 22. The soldier suffers tremendously and gains very little. (7.6, 1.0, 14)
( ) 23. The evils that war brings in its train far outweigh any possible benefits. (8.9, 0.9, 17)
( ) 24. We should have a moderate amount of military training in our schools. (4.4, 0.6, 8)
( ) 25. No scheme of aggression or conquest can be pursued for any considerable length of time without enfeebling victor as well as vanquished. (7.4, 1.1, 13)
( ) 26. When war threatens we should refuse the call to service and increase our anti-war activity. (10.5, 0.7, 20)
( ) 27. It is foolish to talk of the abolition of war, since pugnacity is an ineradicable human instinct. (2.5, 1.0, 4)
( ) 28. There is no progress without war. (0.6, 0.7, 0)
( ) 29. Militarism should be abolished from the curriculum of the state schools. (8.4, 1.1, 16)
( ) 30. It is not in war but in peace and prosperity that our worst vices develop and grow rank. (2.1, 1.1, 3)
( ) 31. We cannot hope to do away with war, because it is part of the unending struggle for survival in a crowded world. (2.9, 1.1, 5)
( ) 32. If armed conflict between individuals and cities can be outlawed, it is possible to outlaw armed conflict between nations. (7.8, 1.0, 14)
( ) 33. Every war shows cowardice, murder, arson, graft, and leaves a trail of personal and national demoralization. (9.5, 1.0, 18)
( ) 34. The most frequent cause of war is the rivalry of nations for possession of territory, markets, concessions, and spheres of influence. (5.6, 0.9, 10)
( ) 35. There in no conceivable justification for war. (10.7, 0.7, 21)
( ) 36. Military training is imperative, but it should be voluntary. (4.1, 0.8, 7)
( ) 37. Nations should agree not to intervene with military force in purely commercial or financial disputes. (6.3, 0.9, 11)
( ) 38. Concerning war we must be abolitionists. (10.0, 1.1, 19)
( ) 39. The abolition of war would mean effeminacy, softness, debilitation, and degeneracy. (1.6, 1.0, 2)
( ) 40. A host of young men entered the war in a spirit of idealism and unselfish devotion to a great cause, only to return disillusioned and cynical as to the value of ideals. (8.1, 1.0, 15)
( ) 41. For the liberty of oppressed nations wars should be fought. (3.4, 1.0, 6)
( ) 42. Compulsory military training should be established in all countries. (1.0, 0.9, 1)
( ) 43. Pugnacity, rivalry and self-interest are natural, but need not result in war any more than human desire for dominance need result in slavery. (6.8, 1.1, 12)
( ) 44. Peace and war are both essential to progress. (5.4, 0.6, 9)

Information

Name.....
Classification (underline one):

Freshman Junior Graduate
Sophomore Senior

Sex (underline one): Male Female
Age:

3. *Scoring.*—For scoring either the original scale values or the equivalent, numbers might be used. In this experiment for the purpose of simplification and because the original scale values were fairly evenly distributed, equivalent numbers were used as a basis for scoring. The score was the average of the equivalent numbers of the statements marked plus.

4. *Reliability.*—For calculation of the reliability of the scale the method of form comparison was chosen. For this purpose the data obtained from four hundred students (see application of scale) were used. To correlate the two forms the product moment coefficient of correlation was used. The correlation was found to be +.83. The Spearman-Brown formula when applied to this gave a value of +.90, which is the expected correlation of the two forms if correlated with two other forms. The latter correlation is to be regarded as the reliability of the total scale as used in its application.

## Application of the Scale

The constructed scale was given to four hundred students attending the University of Chicago. The four hundred students included eight freshmen, one hundred thirty-six sophomores, ninety-six juniors, sixty-six seniors, eighty-three graduates, and eleven unclassified students. The ages of these subjects ranged between seventeen and forty-four, the average being 21.8. Scores on the scale were studied in relation to four factors: Education of students, scholarship, sex, and church affiliation.

1. *Education of Students.*—The four hundred records were arranged on the basis of the educational status of each student. Records of freshmen were put in one group, records of sophomores in another, and so on. All in all, five groups were obtained. The types of groups and the number of students in each are shown in Table III. For each of the groups an average of the scores was then calculated, also the

standard error of the average, and the standard deviation of the distribution of the scores.

A tendency may be observed in the successive differences between the averages of the five groups. If taken separately they are hardly of any significance, but when compared one can observe a certain

TABLE III

| No. | Class | *N* | Mean score | $\sigma_m$ | $\sigma$ |
|---|---|---|---|---|---|
| I | Freshmen | 8 | 11.23 | .71 | 2.02 |
| II | Sophomores | 136 | 11.32 | .20 | 2.35 |
| III | Juniors | 96 | 11.56 | .20 | 2.05 |
| IV | Seniors | 66 | 11.93 | .28 | 2.30 |
| V | Graduates | 83 | 12.82 | .23 | 2.16 |

definite continuity from one group to another. These differences all run in the same direction, which indicates a tendency toward greater pacifism as education increases. This increase is graphically represented in Fig. 3.

The second feature of the differences is their gradual increase. This function, somewhat magnified, may also be seen in Fig. 3. A curve drawn through the averages appears to be postively accelerated, the scores ranging from the mean score of the freshmen class to the mean score of the graduates.

These results tend to indicate that education has something to do with the formation and modification of attitudes toward war and peace. With educational advancement students tend to become more favorable to a peaceful way of solving international differences. This general trend is seen in Table III and Fig. 3, according to which there is a continuous change in attitude toward pacifism as the students advance in the college to the graduate school.

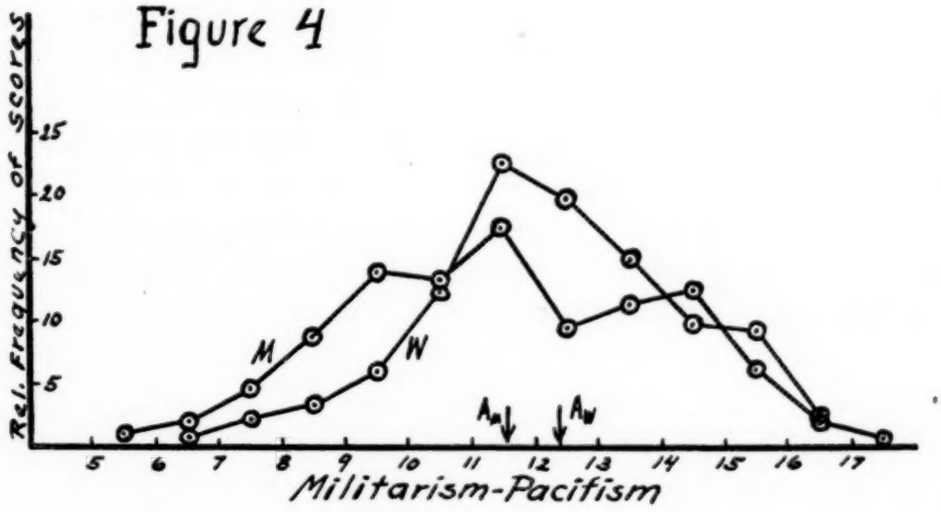
2. *Scholarship.*—Scholarship standings were obtained from the Bureau of Records for two hundred twenty-seven students. The undergraduate grades were expressed by the number of points. But the graduate grades had to be converted into numerical values by the experimenter. This was done as follows: $H = 6$, $P = 4$, $U = 1$, $Inc = 0$. An average score was then calculated for each individual. A range of .50 to 5.50 in scholarship values was obtained. The

Pearson coefficient of correlation between the scholarship values and militarism-pacifism was found to be +.15.

This low correlation seems to indicate that there is little or no relation between scholarship and militarism-pacifism. If there is any, then there is a slight tendency for pacifism to correlate positively with scholarship. Students of higher standing in scholarship tend to be pacifists.

3. *Sex.*—Among the students tested two hundred twenty-five were men and one hundred seventy-five women. The average score for men was found to be 11.54, and for women 12.30, the difference between the two being .76. The standard error of the difference was .20, which is less than a third of the difference. This indicates that women, on the whole, are more pacifistic than the men. The extent to which the two groups overlap may be seen by comparing the two distributions in Fig. 4.

Figure 4




It is also evident from the figure that the distribution of the scores of men is wider than that of women. In terms of the standard deviations the variability was found to be 2.36 for men and 1.98 for women. Women appear to be in better agreement regarding the issue than men, and there appear to be many more militarists among men than among women.

4. *Church Affiliation.*—Data on church affiliation were obtained from the university files. Ten denominations having the largest representation were selected for study as shown in Table IV. For each an average score was calculated, and the denominations were then arranged in descending order.

TABLE IV

| Denomination | Cases | Mean score |
|---|---|---|
| Disciples of Christ | 9 | 13.65 |
| Baptists | 11 | 12.57 |
| Jews | 33 | 12.41 |
| Presbyterians | 24 | 11.95 |
| Congregationalists | 11 | 11.92 |
| Methodists | 24 | 11.49 |
| Christian Scientists | 8 | 11.42 |
| Catholics | 14 | 11.08 |
| Lutherans | 7 | 10.65 |
| Episcopalians | 14 | 10.00 |

The number of cases in each denominational group is not large enough to warrant any definite conclusions, but the results are interesting and suggestive. The most pacifistic churches may be found at the top of the list and the least pacifistic at the bottom. The three churches most similar in form, the Catholic, Lutheran, and the Episcopalian, seem to be similar in attitude and occupy the lowest three positions. That is to say, they tend to be the most militaristic denominations of all the ten. Disciples of Christ, Baptists, and Jews stand at the top of the rank order; Presbyterians, Congregationalists, Methodists, and Christian Scientists are in the middle.

## CONCLUSIONS

The following brief conclusions may be drawn from the experiment described in this article:

1. By the use of the method of equal appearing intervals it was possible to construct an evenly graduated scale of militarism-pacifism. The unit of measurement was one-eleventh of the range of attitude between the most extreme degree of militarism and the most extreme degree of pacifism represented in our list of statements.

2. The reliability of the scale was found to be +.90.

3. Education appears to develop pacifism as measured by the scale. A consistent change toward pacifism was found for the successive undergraduate classes and the graduate classes.

4. A coefficient of correlation of +.15 was found between militarism-pacifism and scholarship.

5. Women appear to be more pacifistic, on the average, than men. Men vary more than women.

6. The Catholic, Lutheran, and Episcopalian seem to be, on the average, the least pacifistic churches of the ten compared. Disciples of Christ, Baptists, and Jews appear to be most favorable to the peaceful method of solving international problems.

# WHAT THE THEORY OF FACTORS IS *NOT*

C. SPEARMAN

University of London

Not long ago there was held a very novel and successful photographic exhibition. The pictures sent in were not as usual the best, but the most instructive. Among them were those got from under-exposures and from over-exposures; those where the immersion in the developer had been too long, and where it had been too short; where the focussing had been faulty, and where the camera had been shaken; where the lens had caused flare, or the bellows had not been light-proof; where the printing paper had been stale, or not of a suitable kind. In a word, the pictures suffered from the thousand and one ills to which the photographic art is heir. The great idea of the exhibitors was, of course, that their own unfortunate experiences should serve to warn others.

Now, if not too presumptuous, I would like to suggest that a somewhat similar general service might be afforded by a book which has recently come from the assembled strength of the psychological department of the University of Minnesota, as represented by Professors Paterson, Elliot, Anderson, Toops, and Heidbreder; a volume, moreover, fortified by the unstinted financial support of the National Research Council.[1]

Taking first the historical aspect of this work, we are told in it that the theory of two factors was not "definitely advanced" till 1914, having previously been only "implied." Now, if such an unhappy account can befall the distinguished Minnesota syndicate, surely there must be many others who are not quite clear on the matter and would gladly be informed. In truth, the very first publication about such factors (this was ten years earlier than the date mentioned above) supplied perfectly definite observations and at the same time a mathematical criterion so exact as to be substantially that used at present. The chief final conclusion of the whole work ran (in italics) as follows:

> All branches of intellectual activity have in common one fundamental function (or group of functions), whereas the remaining or specific elements of the activity seem in every case to be wholly different from that in all the other cases.[2]

Anything more precise I, for my part, could hardly formulate to this day. Further, not only were these two factors, $g$ and $s$,

explicitly proclaimed; even their very influences were actually discovered and measured. What is still more, this original publication already drew the practical corollary (which was so brilliantly exploited the following year by Binet and Simon) that the *g* can actually be measured; this is done by throwing together any hotchpot of tests so numerous and heterogeneous that, on the whole, the influences of the *s*'s more or less perfectly neutralize each other. This, vital corollary was indicated by the very title of the article, "'General Intelligence' Objectively Determined and Measured."

Incomparably more important, however, than any such mere history of the doctrine of two factors is the exposition of its essential nature. According to the authors at Minnesota, the basis of the doctrine is "simple enough"; their characterization runs as follows:

> It is commonly assumed that a perfect hierarchy creates strong presumption that a general factor is at work.

For ventilating this view we must be more grateful to them than ever; since it not only departs from the actual state of affairs, but turns it upside down and inside out. In truth, no "assumption" or "presumption" has ever been made at all about the occurrence, or even measurability, of the *g* and *s*. Instead, rigorous proof is given that in the said case of hierarchy (more strictly speaking, "equi-proportionality"), and solely in this case, the scores in the tests or other variables can *always* be divided into these two factors.[3]

What is the good of this, it may be asked? For can not the tests equally well be divided after any other fashion? The answer is to concede that, so far as the mathematical theorem goes, other modes of division—perhaps an infinite number of them—certainly are possible. But yet others, and extremely important ones, are *not* possible. Thus the tests cannot possibly (where the criterion is satisfied) be so divided as to admit of any group factors (of appreciable size). Again, the tests can not possibly be so divided as to have any general factor other than our *g*. Here we notice that, although the theorem does not in itself supply any evidence that the *g* deserves the title of "intelligence," still it does disprove the existence of any other and different general intelligence. Of course, there may be other mental operations that anyone may like to call "intellectual." But then these can, at most, only constitute classes. They can not possibly be general in the sense of having functional unity.

The further question still remains open, however, as to why we should give preference to the division into *g* and *s* over any other of

the modes of division that the said mathematical theorem really does leave open. A facile reply (but quite correct, so far as it goes) is that this mode of division is the simplest. But penetrating more deeply into the essential character of empirical science, we may answer with greater breadth that these two factors have hitherto shown themselves to be those divisions about which the most important statements and discoveries can be made. To mention only a few—we shall see many more later on—it was in *g* that the inspiration really arose for all the current tests of "general ability," or even of "mental age"; only in *g* do all these tests to this hour find any stable anchorage amid the shifting sands of the arbitary and green-table conceptions of "intelligence"; and with respect to our factors alone can be measured the scientifically indispensable "probable errors."

Another object lesson is afforded by these authors at Minnesota when they embark on the troubled seas of controversy. But here, to speak with frankness, they give some occasion for surprise. They bring against the theory of two factors the following pair of objections. The one is that the inter-correlations between their tests have turned out to be low. Surely we might have expected much less eminent and careful authorities than these to know at least that the theory is in no way based on the absolute magnitudes of the correlations, be these high or low, but solely on their ratios to each other. The second objection raised is that in their results the criterion of "intercolumnar correlations" did not average +1.00. But here their luck was out altogether. A look at the work which introduced and demonstrated this criterion will at once show that they have used an entirely wrong formula;[4] that, moreover, they have applied this under conditions that would render even the right formula quite inapplicable;[5] and yet further that the whole criterion was from the very beginning declared to be only provisional and has now become quite obsolete.[6]

To crown all, even if they had applied the proper criterion and had obtained similar results from this also (as they would actually have done, for in this case the many wrongs end by making a right) even then these results would have been no contradiction to our work. For in this particular case of theirs, our work shows that the criterion ought *not* to be satisfied.

This brings us from the preceding negations of the Minnesota investigators to their positive contributions. Here the lesson at issue is the delicate one of distinguishing between *meum* and *tuum*. In the first place they lay claim to have advanced the general hypothesis

of objectively determined "unique traits"; whereas it is hard to see that these are anything more than the old factors camouflaged under a new name. And in the second place (and it is the theme of the whole book) they represent themselves as having discovered the "unique trait" of "mechanical ability"; whereas in truth the latter had already been discovered (and with what I am afraid was far better evidence and more penetrating analysis) by Dr. J. Cox, who moreover succeeded in so doing by basing his research on the theory of two factors.[7]

Perhaps a clue to the last mentioned and gravest lesson supplied by these authors is to be found in their bibliography. Most concerned here is a book of the present writer, "The Abilities of Man." As was said in its foreword:

> This work is the product of many hands and much patience. The lines of investigation were suggested—and even extensive beginnings made to follow them up—over twenty years ago. Since, there have been carried out a long train of laborious researches, each bringing, as it were, a single stone upon a pre-conceived plan. And here, in this volume, at last, every stone is fitted into its place to build up the common edifice.

Among other things, the book contained a full account of this very mechanical ability at issue. Nevertheless, although this book is actually quoted in the bibliography from Minnesota yet its most relevant portion is wholly overlooked. The authors insert there the oddest lot of the same writer's minor controversial articles.

Now such action of theirs would seem to express a view that is held very widely indeed, so that once more we must be grateful for the present opportunity to comment upon it. This view is to the effect that the work of our school has consisted in arguing on behalf of some highly controversial theorem, and that the only concern of other people with it is to join in the fray, for or against.

In truth, the work of this school may be trenchantly divided into three portions. The one which should be taken first is made up of theorems which are no longer seriously contested by anybody. There is the already mentioned purely mathematical demonstration that, when the tetrad differences are zero, then (and only then) the scores of every individual in every test can be divided into $g$ and $s$ where these are independent of each other as also of all the other $s$'s. Then comes the determination of the probable errors, not only of all these but also of the tetrad differences. For we require to know how large their differences ought to be when, as always happens, the data available are not the "true" correlations, but only values more or less

affected by the errors of sampling. Another theorem, this time obtained not from mathematics but from observation, is that under *some* conditions at any rate the observed tetrad differences do show themselves to have just this magnitude required for the two factors. So much for the portion of the work that should be taken first.

The portion that should, on the contrary, be taken last deals with the attempts to arrive at *ultimate explanations;* it includes, for example, the hypothesis that *g* represents a general "energy." Here, indeed, are problems that will long—perhaps always—be infected more than enough with "presumptions and assumptions." Nothing but trouble can be expected from even approaching them until after arraying and considering all the available relevant facts.

This brings us to the remaining and much the largest part of our work; the one which, accordingly, occupies the great bulk of "The Abilities of Man," constituting its second part, and entitled The Fundamental Facts. Here, as mentioned above, are recorded an immense mass of observations untiringly accumulated by very numerous investigators during a great number of years. One large share of these observations is about the conditions under which the criterion of tetrad differences is or is not satisfied. Already in 1906 the discovery had been made that such satisfaction was liable to be disturbed by what were called "broad" group factors.[8]

These too have been made to yield both measurements and the probable errors of these. Among the more important that have (by means of the probable errors) been shown to be significant are the logical, psychological, verbal, arithmetical, and inventive abilities, as also just the mechanical ability now being put forward at Minnesota as new.

Much broader still, but no longer simply an ability, something rather that depends on ratios between abilities, have come the factors *o* and the still more wonderful *p* (sometimes written as *c*). Hardly less important have been the negative results obtained in this way; the evidence has been that the great majority of the commonly alleged unitary abilities, types, profiles, and so forth are not unitary at all.

All such factors are originally determined in an almost purely statistical manner and therefore are psychologically nearly meaningless. But then come observations that do infuse them with meaning. The relations of all of them have been ascertained to all the qualitative laws of cognition; and in this way the "noegenetic" nature of *g* has been settled, not by biological or philosophical cogitation, but by

actual and detailed measurements of its magnitudes and correlations under varying conditions. At the same time, investigation has been made of the relation of all these factors to the quantitative laws; to those of span, of retentivity, of fatigue, and of conation, as also to purely physiological influences. Furthermore, from all this systematic examination of the sphere of cognition, a passage has been found over to the other great sphere of the mind, that of volition, emotion, and character. Here too a general factor has been discovered, one nearly or quite independent of *g*, and perhaps even more important still.

These observations are, indeed, so presented as to throw light upon the factors involved, general, specific, and group. But they stand equally open to being envisaged from any other angle. They do not in the least depend on any such hypotheses as that of "energy," but on the contrary supply the very facts upon which such hypotheses ought to be accepted or rejected. These observations should not, I like to think, be overlooked by any one interested in any sphere at any rate of individual differences. They cannot legitimately, I feel sure, be disregarded by those who lay claim to original discoveries. Least of all ought they to be ignored by those who make bold to enter into the lists of controversy.

## References and Notes

1. "Minnesota Mechanical Ability Tests." 1930.
2. "General Intelligence Objectively Determined and Measured." *American Journal of Psychology*, 1904, p. 284.
3. Positive proof will be found for instance, on pp. iii–vi of "The Abilities of Man" by the present writer, 1927. This demonstration consists throughout of straightforward mathematics; there is not an assumption or presumption from end to end. For those who find the mathematics in it difficult, it may be mentioned that a much simpler version of it has been devised and will shortly be published (in "Nature") by the well known mathematician, Professor Piaggio.
4. *British Journal of Psychology*, Vol. V, 1912, p. 82, formula (8). The simpler formula employed by them taken no account of the dominant influence of the sampling errors of the correlation.
5. *Ibid.*: P. 56, the lines in italics.
6. "The Abilities of Man." By the present writer. 1927, pp. 138–140.
7. "Mechanical Aptitude." J. W. Cox, 1928, Methuen and Co.
8. Die Korrelation zwischen verschiedenen geistigen Leistungsfähigkeiten. *Zeitschrift fur Psychologie*, Vol. XLIV, 1906, p. 103.

# A REPLY TO SOME RECENT CRITICISMS BY PROFESSOR SPEARMAN

H. D. CARTER

University of Minnesota

Since scientific controversies frequently lead to clarification of issues, a comment on Professor Spearman's recent criticisms[1] of my attempt to analyze the nature of mechanical ability[2] may be in order.

His main criticism was directed against my use of the intercolumnar correlation method as a criterion for the presence of the familiar "two factors," *g* and *s*. He characterizes this method as obsolete, yet in his book, "The Abilities of Man," on page 139, he has a table of twenty series of data which were analyzed by this method, and the results were used in support of his theory. This suggests that even though the method may now be labelled as "obsolete," nevertheless it is permissible to continue to report results secured by that method. Furthermore, it is to be remembered that the newer tetrad-difference method of analysis mathematically expresses the same fundamental relations. In my own work I recognized this fact and accordingly analyzed my data by means of both methods, discovering that the trend of results was the same. This was inevitable in the very nature of things. Since I employed the tetrad difference method in addition to the intercolumnar correlation method, I do not believe that Professor Spearman was justified in his main criticism.

Mr. Spearman also criticized my work by stating that my tables of correlations did not satisfy the only condition under which the intercolumnar criterion is applicable. It is true he has a criterion for the exclusion of certain pairs of columns, which is arbitrary, and not wholly free from criticism. After reading Thompson's mathematical criticisms[3] of the "exclusion criterion" I did not regard this criterion as essential, but I decided, nevertheless, to apply it to my main table of intercorrelations. It was my misfortune, however, to take this criterion directly from Professor Spearman's 1914 paper,[4]

---

[1] Spearman, C.: A Truce to "Barking in." *Journal of Educational Psychology*, Vol. XXI, 1930, pp. 110–111.

[2] Carter, H. D.: The Organization of Mechanical Intelligence. *Journal of Genetic Psychology*, Vol. XXXV, 1928, pp. 270–285.

[3] Thomson, G. H.: On the Degree of Perfection of Hierarchical Order among Correlation Coefficients. *Biometrika*, Vol. XII, 1919, pp. 355–366.

[4] Spearman, C.: The Theory of Two Factors. *Psychological Review*, Vol. XXI, 1914, pp. 101–115. (See page 112.)

in which he himself misstated it. I should have secured the method as correctly stated by Spearman in his 1912 paper.[1] The criterion, therefore, was improperly applied, which led to my incorrect statement that all columns of my Table III satisfied it. But recently, after reading his criticism, I applied the criterion again, this time correctly. Of the eight columns in Table III, four satisfied it, and four did not. The average correlation between all columns in the table was .41. The average of the six intercorrelations of the columns which did satisfy the criterion was .39. For all other pairs of columns the average intercorrelation was .42. Correct application of the criterion is thus shown to have no important bearing on the results.

Professor Spearman also criticized my paper because his own references to one published article and to two unpublished reports in his "The Abilities of Man" were neglected. He claims that those investigations reveal mechanical ability to be unrelated to *g* and to involve a group factor. But my analysis showed overlapping group factors. It is to be noted that my results differ from those quoted by Spearman in emphasizing group factors rather than a group factor. Thus it would appear that the Minnesota Mechanical Research Project makes an additional contribution to previous knowledge dealing with the abilities of man.[2]

---

[1] Spearman, C. and Hart, B.: General Ability, Its Existence and Nature. *British Journal of Psychology*, Vol. V, 1912, pp. 51–84. (See page 56.)

[2] Paterson, D. G., Elliott, R. M., Anderson, L. Dewey, Toops, H. A., and Heidbreder, E.: "The Minnesota Mechanical Ability Tests." University of Minnesota Press, 1930, pp. 1–640. See especially Chapter II, The Theory of Unique Traits; Chapter XI, The Organization of Mechanical Ability; Chapter XII, Mechanical Ability as a Unique Trait.

# STUDIES IN HANDEDNESS: III. RELATION OF HANDEDNESS TO SPEECH

R. H. OJEMANN

State University of Iowa

In this article data will be presented relative to the problem of the effect produced upon the speech function by training left-handed children to write with the right hand. For this investigation two groups of subjects were used. In the unselected group of five hundred eighteen pupils used in the investigation reported in the first article of this series of studies in handedness there were twenty-seven pupils who were diagnosed to be left-handed by the technique described in the first paper. Seven of the twenty-seven cases wrote with the right hand. These seven cases constitute the first group of dextrosinistrals[1] used in the present investigation.

The second group of dextrosinistrals was obtained as follows: In the city from which the random sample of five hundred eighteen pupils was taken from Grades III–VIII there was a total number of 1571 pupils in these grades. It was assumed that if the names of all the pupils who performed one or more of such activities as using a pair of scissors, throwing a ball, and dealing cards were obtained, several dextrosinistral cases would be secured. Accordingly, each teacher of Grades III–VIII was asked to make three lists of names: (1) A list of the names of all the pupils who used a pair of scissors with the left hand, (2) a list of the names of all the pupils who threw balls with the left hand, and (3) a list of the names of the pupils who thought they could deal cards better with the left hand than with the right hand. The purpose of this request was made clear to each teacher. It was pointed out that it was desirable to get all the pupils who used the left hand for the activities mentioned; and that, therefore, the teacher should include all doubtful cases. It was explained that it would make no difference if too many cases were reported, since each case would be carefully tested. It was also explained that it would be no reflection upon the teacher's ability if too many cases were reported. Each teacher responded enthusiastically. From the list of names submitted by the teachers the names of those students

[1] This term was adopted by Ballard: Sinistrality and Speech, *Journal of Experimental Pedagogy*, Vol. I, 1911–1912, pp. 298–310, and will be used in this report to designate a left-handed individual who has learned to write with the right hand.

who had been included in the unselected group were checked. The pupils whose names remained were carefully tested. By the use of this procedure sixteen cases[1] were secured in which the combined score on the five unimanual handedness tests was negative four or lower and in which the right hand was used for writing. These sixteen cases constitute the second group of dextrosinistrals used in the present investigation.

In addition to the unimanual and bimanual handedness tests each dextrosinistral was given three speech tests:

1. *Articulation Test.*—This test consisted of a series of sentences which were read aloud by the subject while the experimenter checked on a record sheet provided for the purpose those sounds that were made incorrectly.

2. *Spontaneous Speech Test.*—The subject was shown a picture and was asked to describe aloud what he saw. The experimenter made detailed notes of any speech defects.

3. *Oral Reading Test.*—Gray's "Oral Reading Check Tests" were used. A record was kept of the time required to read the whole selection rather than each paragraph. This modification was made in order that the experimenter might direct his undivided attention to the errors made, since an accurate record of the errors rather than of the rate of reading was desired.

In addition to the data gathered by means of the handedness and speech tests, the parents were interviewed to obtain their judgments concerning the handedness of the subjects and to obtain the facts relative to the history of the subjects' speech and writing habits.

The data for the two groups of dextrosinistrals, twenty-three cases in all, are presented in Table I. The first seven cases are the dextrosinistrals of the unselected group of five hundred eighteen pupils. The cases in each group are arranged according to the grade reached in school at the time this investigation was made. Since the investigation was made shortly after the second semester had begun, the age of the subject at the beginning of the second half of the school year was used for the data in the column headed "Age." The unimanual test scores are given in terms of unit scores. (For a definition of unit scores see the first paper of this series.)

[1] The proportion of dextrosinistrals in the unselected group is .0135 ± .003. Since sixteen cases were found in the remaining group of pupils, the proportion is .0152 ± .003. The difference between the two proportions is .0017, which is less than the probable error of either proportion.

TABLE I.—DEXTROSINISTRALS STUDIED IN THIS INVESTIGATION

| Subject number | Sex | Grade | Age | Unimanual test scores (unit scores) | | | | | | Bimanual test scores | | Speech test results | Parents' judgments of handedness |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | Cut.[2] | Tap.[3] | Thr.[4] | B-P.[5] | N-T.[6] | Combined | Sw.,[7] etc. | Bat. | | |
| 494 | G | 3 | 9-6[1] | −5 | −1 | 3 | −1 | −3 | − 7 | L | L | No speech defects | Left-handed |
| 184 | B | 5 | 10-8 | 0 | −1 | −3 | −2 | −2 | − 8 | R | L | No speech defects | Left-handed |
| 407 | B | 5 | 10-2 | −3 | 0 | −3 | 0 | −3 | − 9 | L | R | Faulty articulation | Left-handed |
| 104 | G | 6 | 11-6 | −2 | 0 | 0 | −2 | 0 | − 4 | L | R | No speech defects | Left-handed |
| 19 | B | 7 | 12-6 | 1 | −1 | −3 | −2 | −2 | − 7 | L | R | No speech defects | Left-handed |
| 271 | B | 8 | 13-5 | −2 | 0 | −3 | −1 | −3 | − 9 | L | L | No speech defects | Left-handed |
| 272 | B | 8 | 13-7 | −1 | 2 | −3 | 0 | −3 | − 5 | L | R | No speech defects | Left-handed |
| 735 | B | 3 | 9-2 | −1 | −1 | −3 | −1 | −1 | − 7 | L | R | No speech defects | Left-handed |
| 784 | G | 3 | 16-6 | −1 | 2 | −3 | −3 | 0 | − 5 | L | R | No speech defects | Left-handed |
| 652 | B | 4 | 9-6 | −2 | −2 | −3 | 0 | −1 | − 8 | L | R | No speech defects | Left-handed |
| 676 | G | 4 | 10-1 | 4 | −1 | −3 | −6 | −6 | −12 | R | R | No speech defects | Left-handed |
| 609 | B | 5 | 10-9 | 0 | −3 | −3 | −2 | −1 | − 9 | L | R | No speech defects | Left-handed |
| 751 | G | 5 | 10-9 | −5 | −1 | −3 | 0 | −2 | −11 | L | L | No speech defects | Left-handed |
| 619 | G | 6 | 11-5 | −3 | 1 | −3 | −2 | −3 | −10 | R | L | No speech defects | Left-handed |
| 636 | G | 6 | 11-3 | −2 | −2 | −3 | −1 | −2 | −10 | L | L | Lisp | Left-handed |
| 703 | G | 6 | 11-2 | 1 | −2 | −3 | −1 | −3 | − 8 | L | L | No speech defects | Left-handed |
| 705 | G | 6 | 11-4 | −1 | 0 | −3 | −1 | 0 | − 5 | L | L | No speech defects | Left-handed |
| 603 | B | 7 | 12-3 | 1 | −1 | −3 | −5 | −3 | −11 | L | L | No speech defects | Left-handed |
| 692 | B | 7 | 13-6 | 0 | −2 | −3 | −2 | −1 | − 8 | L | L | No speech defects | Left-handed |
| 708 | B | 7 | 12-6 | 0 | 3 | −3 | −1 | −6 | − 7 | R | R | No speech defects | Left-handed |
| 748 | B | 7 | 12-7 | −4 | 1 | −3 | 0 | −1 | − 7 | L | R | No speech defects | Left-handed |
| 760 | G | 7 | 13-2 | −2 | −1 | −3 | −2 | −2 | −10 | L | R | No speech defects | Left-handed |
| 773 | G | 8 | 13-3 | −1 | −1 | −3 | 0 | 0 | − 5 | L | R | No speech defects | Left-handed |

[1] To be read, nine years and 6 months.
[2] Cutting.
[3] Tapping.
[4] Throwing.
[5] Block-packing.
[6] Needle-threading.
[7] Combined score on sweeping, raking, and shoveling.

A study of the ages of the subjects and the grades reached in school shows that one of the subjects—Subject 784—was seriously retarded in her school work. This subject had an IQ of sixty-two[1] and was afflicted with a mild epileptic tendency. According to her third grade teacher, these facts accounted for her extreme retardation.

A study of the combined unit scores on the unimanual tests shows that all the cases scored negative four or lower and are, therefore, definitely left-handed. This diagnosis was corroborated in every case by the parent's judgment.

A study of the speech test results shows that in two cases, Subject 407 and Subject 636, speech defects were present at the time of the investigation. Subject 407 failed to articulate accurately the "b," "p," and "lz" sounds. His difficulty was diagnosed as faulty articulation. This subject showed no tendency to stutter. Subject 636 had a mild lisp. She substituted the "th" sound for the "s" sound. No additional speech disturbances were detected by the spontaneous speech test or by the oral reading test.

The results of the speech tests were confirmed by the interviews with the parents. According to the parents' reports only Subjects 407 and 636 had speech defects at the time of the investigation.

Subject 104, at about two years of age, when she first began to talk, stuttered noticeably. This speech disturbance continued in mild form until she reached the third grade. The disturbance then disappeared. No speech defects were present at the time this investigation was made. The first grade teacher attempted to teach the subject to use her right hand for writing and was successful without much difficulty. The mother stated that this was due to the child's temperament. The child was easily managed and the teacher, therefore, had little difficulty in starting the child to practice the use of the right hand in writing. Since the first grade, the use of the right hand for writing had been consistent. The parents reported that this training did not increase the severity of the stuttering.

Subject 19 developed a very marked case of stuttering when he was about two years of age. This disturbance lasted for about a year, after which it gradually disappeared. He had developed no speech disturbances since that time. During the pre-school period the subject used the left hand for writing. Under the influence of

---

[1] The writer is indebted to the school authorities for this information. The Binet-Simon test was used.

the first grade teacher, who treated the child's difficulties sympathetically, the subject learned to write with the right hand. No effect upon the speech function of the training in the use of the right hand for writing was observed by the parents.

Subject 676 stuttered slightly when she first learned to talk. This disturbance had disappeared by the time the child was three years of age. No speech disturbances had developed since that time. The subject used her left hand for writing until she entered the second grade. Her second grade teacher persuaded her to practice writing with the right hand. The teacher encouraged the child and exercised considerable patience in helping her overcome the difficulties. Since the second grade, the subject has used the right hand for writing.

When Subject 703 was about three years of age the mother attempted to train her to perform the principal unimanual activities right-handed. The child began to stutter. The mother became alarmed and, on the advice of a neighbor, discontinued the training. The stuttering disappeared about a year and a half later and did not appear again. When the child entered the first grade the parents insisted that the teacher train the child to write with the right hand. The teacher and the parents treated the child's difficulties sympathetically. No effect upon the speech function of the training in writing with the right hand was observed.

The foregoing data may be summarized thus: Two of the twenty-three dextrosinistrals evidenced speech defects at the time they were tested. In these two cases the speech disturbances appeared when the child first learned to talk. Four of the twenty-three subjects were reported to have had a speech disturbance at some time previous to the investigation. In these cases no connection between the training in using the right hand for writing and the speech disturbance could be established. The remaining seventeen dextrosinistrals had developed no speech defects at any time.

Before an interpretation is made of the above data two additional facts relative to the history of the subjects' writing habits must be noted. In all twenty-three cases the intensive training in the use of the right hand was started after the speech habits had become well organized, that is, after the subject had entered the first grade. In two cases punishment was used by the teacher to train the child to write with the right hand. The second grade teacher reprimanded Subject 619 severely and punished her by slapping the hands, shaking, etc. The child had considerable difficulty in learning to write with

the right hand. Her early attempts produced writing in mirror form. The treatment she received by her second grade teacher made her nervous and irritable but did not, according to the parent's report, affect her speech. Subject 692 was punished by slapping the hands and by shaking. His first grade teacher frequently strapped the left hand to the body to prevent the subject from using it during the writing period. This treatment made the child nervous but did not produce a speech disturbance.

The data presented in this paper are not subject to any serious limitations. In all twenty-three cases there is no doubt concerning the handedness of the subjects or concerning the speech disturbances. The data form an impressive array of facts the interpretation of which will be briefly considered.

These data tend to show that in training a left-handed individual to write with the right hand, the handedness of the subject is not a sufficient condition to produce a speech disturbance. It appears to be the exception rather than the rule for a speech disturbance to be produced by training left-handed individuals to write with the right hand. Any explanation for those cases in which a speech disturbance follows the shift from the left hand to the right hand must therefore be based upon a factor or group of factors which operates only in exceptional cases. An explanation which is based solely upon the neurological basis of left-handedness is therefore inadequate. The nervous system is far more flexible than is implied in such theories. The various motor and speech areas in the cortex are not localized to such an extent that it is impossible for a left-handed individual to learn to write with the right hand without disturbing the speech coordinations. Most left-handed individuals who learn to write with the right hand do not develop speech disturbances.

It is, of course, possible that in training a left-handed child to write with the right hand, factors other than the conflict between the right and left hand may be brought into play to produce a speech disturbance. For example, if the subject is emotionally unstable and is given no special care, or if the teacher employs a method of training that produces a serious mental confusion on the part of the subject, the probability of producing a speech disturbance is much greater than if an attempt is made to treat the child's difficulties intelligently.

The data presented in this study show, however, that under ordinary conditions the danger of producing a speech disturbance, after the speech habits have been formed, by training a left-handed child to

write with the right hand is very slight. The nervous system is sufficiently flexible to allow an individual having a strong left-handed tendency to be trained to develop the complex coordinations required for writing with the right hand without bringing about a disturbance in a closely related series of fine coordinations such as are involved in speech.

# THE CONDITIONING OF OVERT EMOTIONAL RESPONSES

HAROLD ELLIS JONES

University of California

The laboratory record which has been chosen for this report derives interest from the fact that it duplicates, in a human subject, and in a single experiment, many of the typical phenomena which have been described in the literature on conditioned reflexes in the lower animals.

The experiment was conducted in an isolated chamber, on a platform inlaid with thin brass strips; the strips were one-quarter inch in width, separated by quarter inch spaces, and connected alternately to the positive and negative primaries of a Porter inductorium. For greater quiet, a fifty dv. tuning fork was substituted for the reed make-break of the inductorium. Care was taken to adjust the secondary coil of the inductorium so as to give an electrotactual impulse restricted to a range of low intensities; the stimulus value may be described as a mildly uncomfortable "tickle" rather than as a "shock." The subject for this experiment was Robert B., a child fifteen months of age, of a markedly stolid and apathetic disposition, able to walk, and possessing a speaking vocabulary limited to fewer than five words. When he was brought to the laboratory a group of toys on the platform attracted his attention. He sat down in the middle of the platform and remained contentedly at play when left alone in this situation. Two observers, concealed behind a one-way vision screen three feet away, recorded the child's overt behavior in as great detail as was possible by a simple observational method. The inductorium, batteries, and a system of bells and buzzers, were housed in an apparatus box in a corner of the room. Robert's original reaction to an electric bell, which was sounded for approximately two seconds, could be described as "mild interest," shortly becoming "indifference." After four repetitions he no longer gave any observable reaction to this auditory stimulus. Stimulations were now given, at intervals of about fifteen seconds, associating the bell with the electrotactual stimulus: After two seconds with the bell, the inductorium circuit was established, and auditory and tactual stimuli were continued together for three seconds. The application of the tactual stimulus (felt either on hand or foot, or both) resulted in a prompt shifting of the skin areas which were in contact with the conducting strips; in each case a

slight but unmistakable startle reaction was registered, accompanied by expressional changes such as a box-shaped mouth, or a puckering or pouting of the lips, and frequently by vocal murmurs or a slight whimpering which ceased at the end of the stimulation period. In the periods between stimulations, Robert played quietly with the toys, and showed no overt indications of a persisting apprehensiveness or emotional upset. After three associations with the primary stimulus, the bell was sounded alone; for the ensuing five stimulations, startle reactions were recorded which were indistinguishable from those elicited by the primary stimulus; the second observer, who of course heard the bell, but had no direct means of knowing when the electrotactual stimulus was used, judged that the associations with the unconditioned stimulus were being continued through eight (instead of through merely the first three) successive trials. Two stimulations by a buzzer were also effective in giving startle reactions, although with appreciably greater delay. A small hand bell, of very different timbre and frequency from the conditioned stimulus, elicited no reaction. In the following trials an extinctive inhibition was established to the bell, and six reinforcements were given; in five of these, the electrotactual stimulus was received on the foot and not on the hand, and it was noted that the stimulus to the foot showed a reduced effectiveness. At the fiftieth trial, after eighteen presentations of the conditioned stimulus alone, a C-R was still present: The subject moved his foot sharply, turned his head away, and vocalized in a manner interpreted as "scolding."

After twenty-four hours, the child was again brought to the laboratory. He entered the chamber readily, seated himself on the platform, and began playing with blocks. At the first presentation of the bell he paused and listened; at the second he started slightly and murmured; at the third he began to whimper, and showed overt movements similar to those which had been elicited at the beginning of the experiment by the primary stimulus. For the buzzer, disturbance was less marked, and for the hand bell only attention responses were recorded. A small piece of *zwieback* provided Robert with an activity which for the time being inhibited overt responses to the bell. In the later stages of extinctive inhibition of the overt response, it was noted that "implicit" responses occasionally still occurred, as illustrated by pupillary dilation or by flushing.[1]

[1] A fuller instrumental study of conditioned reactions has been reported in Jones, H. E.: The Retention of Conditioned Emotional Reactions in Infancy. *Journal of Genetic Psychology*, Vol. XXXVII, 1930, pp. 485–498.

After seventy-two hours, the first application of the bell resulted in momentary crying; on the second he cried and withdrew his hand slowly; on the third he showed a generalized bodily startle reaction. These were the most marked responses which were obtained during the entire series, and the experiment was discontinued in order to avoid further emotional disturbance. It should be noted that the conditioning was distinctly to the bell and not to the platform; in the intervals between stimulations he played about freely in contact with the brass strips, and showed no hesitation in touching them. There was no conditioning against the total situation (of entering and playing in the room), and no indication that the experiment produced any harmful carry-over in the child's normal play activities.

Opportunities were lacking to test the later course of the C-R; in view of the mildness of the primary stimuli employed in the original conditioning, it may be expected that the C-R would be short-lived; the effect of telephone and other bells would be to inhibit rather than to maintain the reaction. In conducting an experiment of this nature, it is obviously desirable to select a child of a stable make-up and to proceed with careful attention to the possibility of psychological damage. At the age under consideration, and with properly supervised conditions, it is fair to say that the experiment involves less risk of emotional upset, or of any persisting ill effects, than an equivalent amount of time spent in a routine physical examination.

For purposes of classroom presentation, the writer has found it useful to summarize the foregoing data in the form of the following table:

| Stimulation number | Stimulus | Overt response | Process |
|---|---|---|---|
| | Current | ++ | Startle reaction to unconditioned stimulus |
| 1–4 | Bell | + | Investigatory reactions |
| 5–10 | Bell | − | Negative adaptation |
| 11–13 | Bell, current | ++ | Conditioning |
| 14–18 | Bell | ++ | Startle reaction to conditioned stimulus |
| 19–20 | Buzzer | + | Generalization of the C-R (irradiation of excitation) |
| 21 | Hand bell | − | Specialization of the C-R (discrimination) |
| 22–24 | Bell | + | |
| 25–26 | Bell | − | Extinctive inhibition |
| 27–32 | Bell, current | + | Reinforcement |
| 33–50 | Bell | + | Re-established C-R |
| After twenty-four hours: | | | |
| 51–54 | Bell | − to + | Summation (?) |
| 55 | Buzzer | + | |
| 56 | Hand bell | − | |
| 57–60 | Bell, food | − | Temporary external inhibition |
| 61–62 | Bell | + | |
| 63–65 | Bell | − | Extinctive inhibition |
| After seventy-two hours: | | | |
| 66–70 | Bell | ++ | Spontaneous recovery (disinhibition) |

# CERTAIN AMBIGUOUS TERMS IN EDUCATIONAL PSYCHOLOGY

STEPHEN MAXWELL COREY

De Pauw University, Greencastle, Indiana

It is difficult to understand why educational psychologists should be so little concerned with the terminological status of their field. Many words, if we may judge by the frequency with which they appear, are intended to be extremely meaningful, but what they mean is vague. Yet textbook writers state dogmatically that a "habit" is this, or "learning" is that, without apparent realization of or interest in the fact that these terms mean one thing to them, and something else to the majority of their readers. Nor does it seem wise to smile at this disagreement, and contend that it is after all just a matter of words. Any science must depend upon language for its development, and the value of a science is largely a function of the precision and accuracy of the words that symbolize its meanings. When words are used loosely, "the mind approaches a condition where practically everything is a thing-um-bob or what-do-you-call-it." From a certain point of view, educational psychology seems to be in some such state. Too many of the technically used words are indefinite. They are awkward tools for thinking and treacherous as well for their ambiguity leads to confusion, and an inability to distinguish between separate factors.

In an attempt to make this criticism more concrete, some of the meanings attached to the words habit, learning, intelligence and instinct are presented. These are four of the oldest, seemingly most revered, and certainly most widely used terms in educational psychology. Volumes have been written concerning each, yet none conveys a precise meaning. Each is understood, after a fashion, and freely interpreted. All writers use the same symbols, but the ideas for which the symbols stand are practically indeterminable.

To illustrate, consider the term "learning." Probably no single, so-called technical word, is more widely used by men in the field of education. There are learning curves, learning periods, the learning process, psychologies of learning, and so on almost indefinitely. To assume, however, that writers on these subjects have in mind a common idea would clearly demonstrate a lack of acquaintanceship with the facts. The word "learning" is a standardized symbol representing

an unstandardized concept. It is defined by educational psychologists in one or more of the five following ways:

1. Learning as growth.
2. Learning as synonymous with memorizing.
3. Learning as representing any more or less permanent change in behavior not due to maturation.
4. Learning as synonymous with education, or approaching a goal.
5. Learning as connection-forming.

Kate Gordon is one who advocates the conception of learning as growth.[1] She writes of the growth of motor capacities and instincts. Learned behavior is behavior that has "grown" to be more elaborate and subtle. Growth is always in the direction of greater complexity, so that all change is not growth, and hence not learning. Learning takes place only in the transition from simple to complex. Judd, Koffka, Bühler speak of the "growth" of the mind, the "growth" of certain capacities and abilities. They are disciples of the developmental type of educational psychology. Everything is growth, a rounding out, "closure."

To another group learning merely means memorizing. Seashore devotes an entire chapter to what he calls the learning process, but he discusses only memory.[2] Pillsbury assumes a similar point of view when he states that "learning is no more than the formation of associations."[3] This conception of learning was quite common among all early psychologists, and was clearly expounded and developed by Ebbinghaus. Recently, however, some writers have attempted to distinguish between memorizing and learning, maintaining that the former is pre-eminently a function of the cerebrum, while learning is palæencephalic.

Some psychologists, in an attempt to render an all-inclusive definition of learning, state that it is comprised of all more or less permanent modifications of behavior, save those which come wholly from maturation.[4] According to this definition, learning takes place when one loses the ability, once possessed, to recite a particular poem. Colvin writes, "The learning process may chiefly be described in its most

[1] Gordon, Kate: "Educational Psychology." P. 32f.

[2] Seashore, C. E.: "Introduction to Psychology." Chap. 13.

[3] Pillsbury, W. B.: "Fundamentals of Psychology." P. 365.

[4] Cameron, E. H.: "Educational Psychology." P. 165. Learning in its broadest sense comprises all modifications of behavior except those which result from purely physiological changes, and those which come wholly from increased maturity.

general terms as the modifications of the reactions of an organism through experience."[1] He does not rule out physiological or maturity changes, provided they be effected or influenced by experience.

To other writers, learning has always had philosophical implications. Claparède, in a section of his book where learning and education are used interchangeably, writes, "It [?] is not merely a matter of exercising the intelligence, of furnishing the memory, but rather of directing the character upright, of stimulating zeal, and of developing the will and personality."[2] This concept of learning as a change of behavior in a certain direction, toward a goal, is rather common. Learning is called "profit from experience,"[3] it requires evaluation, and therefore cannot be studied solely with the methods at the disposal of psychologists. How can psychology decide whether a certain change in behavior represents profit or loss?

Again, learning is not only considered as a modification of behavior, but also as a modification of neural conditions. In the face of our relative ignorance of its nature, many writers in education and psychology define learning in terms of the changes in the nervous system which accompany it. Thorndike's statement that "learning is connection forming" carried weight. Starch reproduces the thought in these words: "Probably all forms of learning can be reduced to one relatively schematic type. Reception of impressions through the senses; assimilation, analysis, and combination of processes in the mind; and redirection of impulses to produce a reaction; or in brief, stimulus, association, and response."[4] Gates writes similarly, "learning consists in the strengthening and weakening of connections between situation and response."[5]

The natural result of this lack of agreement as to the nature of learning is apparent. The interested and intelligent layman thinks that educators don't know what they are talking about, and if they cannot agree among themselves as to what learning is, how can they teach his child? His reasoning is not easily criticized, and the lack of agreement is all the more confusing because it is not merely one of words, but involves ideas as well.

---

[1] Colvin, S. S.: "The Learning Process." P. 1.

[2] Claparède, Ed.: "Experimental Pedagogy." P. 58.

[3] Griffith, Coleman R.: "General Introduction to Psychology." P. 119. In other words, true profit from experience (learning) must represent a functional change in the nervous system . . .

[4] Starch, Daniel: "Educational Psychology" (Revised Edition). P. 127.

[5] Gates, A. I.: "Psychology for Students of Education." P. 238.

The concept of "habit" is in much the same predicament. There are almost as many different ideas as to the nature of habit as there are outstanding educators. Some attach a different meaning to the word each time it is used. Obviously, like learning, habit formation may be and is considered from two basically different points of view, neurological or behavioristic. Many authors make this distinction clearly, but in their subsequent use of the term, no hint is given as to which side of the coin is being considered. Warren defines habit as both a "process of forming connections in the nervous arc"[1] and as "an individually acquired and stereotyped series of responses or thoughts."[2] Is each of these a habit? Is a habit only present when both conditions are fulfilled? Do the two occur concomitantly, or does it make any difference?

More pertinent to the subject, even among those authors who attain and consistently hold one point view, neurological or behavioristic, there is a lack of unanimity as to what a habit is. Some, for example Judd who speaks of "Learning or habit formation,"[3] consider the term as equivalent to learning. Whatever is learned is therefore habitual. On the other hand, habit is thought of as something which is largely automatic, mechanical, stereotyped, less dependent upon consciousness. Pillsbury defines habit as a tendency, "the tendency of all movements and of mental operations which involve little or no movement to become mechanized."[4] Or, in the words of James, "An acquired habit is nothing more than a new pathway of discharge formed in the brain, by which certain incoming currents *ever after* tend to escape."[5] The main concern of this group is the determination of a basis upon which to distinguish habits from instincts.

Nor have the possibilities for a misinterpretation of the word "habit" been exhausted. Many writers cannot use the term without involving ethical considerations. He is habitually neat, or honest, or cowardly, or courteous; he has "good" or "bad" habits. According to the many followers of Thorndike a habit is a specific, automatic response to a definite situation, but the extent to which we are said to have the habit of honesty is a function of the number of *widely different* situations to which we respond in a *generally* honest manner.

---

[1] Warren, Howard C.: "Elements of Human Psychology." P. 253.
[2] *Ibid:* P. 403.
[3] Judd, C. H.: "Psychology of Secondary Education." P. 54.
[4] Pillsbury, W. B.: "Education as the Psychologist Sees It." P. 90.
[5] James, William: "Psychology." Briefer Course. P. 134.

There are racial habits, specific habits, general habits, mental habits, and moral habits; in terms of the nervous system or in terms of behavior. To one a habit may be anything learned, to another only those reactions which are stereotyped. Probably an author is not certain just what he means himself; he may rest assured that his own meaning is seldom the one conveyed to the reader.

The concept of "instinct" is so confusing that many authorities advocate abandonment of the term in anything approaching a scientific discussion. This ambiguity is not so much a matter of the existence of instincts as it is a conflict over the best way to explain them and which reactions should be labeled instinctive. Instincts, too, may be considered from either a neurological or behavioristic point of view. All writers stress their inherited nature, but there is little agreement as to their purposiveness.[1] Some attempt to differentiate clearly between reflexes and instincts,[2] while others state that any distinction is based upon relative complexity and is made for convenience only.[3] Instincts are classified in terms of neural organization, behavior, the situations stimulating them, the purposes which they serve, or the emotions accompanying them. Some writers list many,[4] others only a few.[5] There is no concensus of opinion as to whether a certain reaction should still be called instinctive after it has been modified by experience, or whether it should be termed learned.[6] In one context an instinct is a specific reaction composed of a number of reflexes;[7] in another it is a quite variable pattern characterized by its purposiveness and progress toward a preconceived or unconceived goal.[8] Some writers think of intelligence and instinct as being antithetical. A typical quotation, though seldom appearing in so naive a form, is: "The behavior of animals is regulated by instinct; that of man by intelligence, by reason." Opposing this view, Myers writes: "I conclude then, that instincts are not, as has been generally

---

[1] Yet McDougall writes: "Outlines of Psychology." P. 71. But the great majority of all parties would agree that we may properly call instinctive those reactions of animals which seem to be purposive . . .

[2] Koffka, Kurt: "The Growth of the Mind." P. 90ff.

[3] Cameron, E. H.: *Op. cit.*, p. 42.

[4] Thorndike, E. L.: "Educational Psychology." Vol. I.

[5] McDougall, William: "Outlines of Psychology." Chap. V.

[6] Morgan, C. Lloyd: "Animal Behavior." P. 71.

[7] Watson, John B.: "Behavior." P. 106. An instinct is a series of chained reflexes.

[8] Koffke, Kurt: *Op. cit.*, Chap. 3.

supposed, identifiable with reflexes; nor are they, as others have urged, a *tertium quid* beside reflexes and intelligence. According to my view and my use of the word, instinct, regarded from within becomes intelligence; intelligence regarded from without becomes instinct."[1] To Myers the difference between instinct and intelligence seems to depend upon the position of the observer.

We are instinctively kind, clever, stubborn, or critical. All that we are is the sum of our instincts and experiences, an educational platitude frequently stated but as meaningless as the statement that the word "it" is nothing more that "i" plus "t." The word "instinct" is used loosely and represents no definite idea. It may mean any one of a half dozen things, yet its use continues in our scientific discussions.

Lastly, the word "intelligence," though readily and freely used in technical writings may represent any one of the following concepts:

1. Ability to adapt oneself to novel situations.[2]
2. Capacity for knowledge plus knowledge possessed.[3]
3. Inhibition, analysis, perserverance.[4]
4. Capacity to acquire capacities.[5]
5. Capacity to learn or profit by experience.[6]
6. Judgment.[7]
7. The functioning of two factors, one general and operating in all situations, the other specific and operating in each particular situation.[8]
8. Abstract thinking.[9]

---

[1] Myers, C. S.: *British Journal of Psychology*, Vol. III, 1910, p. 218.

[2] Pintner, R.: *Journal of Educational Psychology*, Vol. XII, 1921, p. 139. I have always thought of intelligence as the ability of the individual to adapt himself adequately to relatively new situations in life.

[3] Henmon, V. A. C.: *Op. cit.*, p. 201f. Intelligence involves two factors, capacity for knowledge plus knowledge possessed.

[4] Thurston, L. L.: *Op. cit.*, p. 201f. Intelligence . . . contains at least three psychologically differentiable components: (1) Capacity to redefine an instinctive adjustment . . . (2) Capacity to inhibit an instinctive adjustment . . . (3) The volitional capacity to realize the modified instinctive behavior into overt behavior . . .

[5] Woodrow, Herbert: *Op. cit.*, p. 208.

[6] Dearborn, W. F.: *Op. cit.*, p. 211.

[7] Binet and Simon: *Année Psychologique*, Vol. XI, 1905, p. 195. There is in intelligence, it seems to us, one fundamental organ . . . this is judgment.

[8] Spearman, C.: "The Nature of Intelligence and the Principles of Cognition." P. 4ff.

[9] Terman, L. M.: *Journal of Educational Psychology*, Vol. XII, 1921, p. 728. An individual is intelligent in proportion as he is able to carry on abstract thinking.

This list could be greatly extended. Some definitions would include moral and personality traits as a part of intelligence; some would include the emotions; some would consider the word as being descriptive of behavior; some as being a static component of the "mind."[1]

The lack of concern over this state of affairs is strikingly illustrated by Dearborn, who begins his share of a symposium upon the nature of intelligence thus, "The commonly accepted definition of intelligence is . . . "[2] The very series of articles to which he was contributing brought into relief the fact that there is no "commonly accepted" definition of intelligence, nor is there much of an idea as to its nature. Pressey comes out with the startling confession that, although he spends most of his time measuring intelligence, he is not much interested in its nature.

This lack of agreement over the nature of intelligence is more than one of definition. Briefly considered there are two extreme points of view held by Thorndike and Spearman respectively, with other authorities occupying the ground in between. Thorndike believes that we have as many intelligences as we face situations, while Spearman believes that his *G* operates in all reactions. Quoting Thorndike: "A table of the known degrees of relationships would abundantly confirm the statement that the mind must be regarded, not as a functional unit, nor even as a collection of a few general faculties which work irrespective of particular material, but rather as a multitude of functions, each of which involves content as well as form, and so is related closely to only a few of its fellows, to the others with greater and greater degrees of remoteness."[3]

Interpreting a similar table of relationships Spearman writes: "And thus there emerges the concept of a hypothetical general and purely qualitative factor underlying all cognitive performances of any kind . . . The factor was taken, pending further information, to consist in something of the nature of an energy or power which serves

---

[1] Yet we run across the following quotation. Monroe, Walter S.: "Directing Learning in the High School." P. 302. Some words that are used with only a general meaning in conversation and in non-technical fields, have been assigned precise meanings in the field of education. Examples of such words are: Standards, intelligence, . . . etc.

[2] Dearborn, W. F.: *Journal of Educational Psychology, op. cit.*, p. 211.

[3] Thorndike, E. L.: *Educational Psychology*, Vol. III, p. 366.

in common the entire cortex (or possibly even the whole nervous system).[1]

All of the available space between Thorndike and Spearman is occupied by authorities. Everyone can find a definition of intelligence to suit his temperament and capacities. As Spearman has so well put it: "The reason is now evident why all search for the meaning of intelligence has, even with the greatest of modern psychologists, always ended in failure. It is simply that, in point of fact, this word in its ordinary present day usage, does not possess any definite meaning."[2]

The fault with educational writers is that too many attempt to give a pedigreed meaning to a mongrel word. Long ago the chemist learned that if he wished to convey a respectable thought he must use a respectable vehicle of expression. He no longer speaks of table salt, but rather of sodium chloride. It is as impossible to extract precise, scientific meanings from words on the tongue of every novelist as it is to inject such meanings into words which already symbolize any number of different concepts.

If popular terms can't be isolated from popular meanings new ones must be manufactured, and lecturers and novelists will not make the change.

---

[1] Spearman, C.: *Op. cit.*, p. 5.

[2] *Ibid.*: P. 19.

# THE AUTOMATIC PREDICTION OF SCHOLASTIC SUCCESS BY USING THE MULTIPLE REGRESSION TECHNIQUE WITH ELECTRIC TABULATING AND ACCOUNTING MACHINES

DAVID SEGEL

Department of Research, Long Beach City Schools

The prediction of scholastic success in general, and in the secondary school and college field in particular, is in need of devices for shortening the labor and time necessary for calculating individual prediction scores. This is true mainly because the prediction of scholastic success usually involves the use of the multiple regression equation. The work required to calculate scores after a multiple regression equation is obtained is almost prohibitive in a practical situation. It is possible, as Hull has shown,[1] to construct special machinery which will make predictions according to regression equations that have been worked out. However, the cost of constructing such machinery for this specialized use probably will prohibit its extended use for some time. In the meantime it seems desirable to adapt the use of automatic calculating machinery more generally available to the problem.

We found that the tabulator and key punch units manufactured by the International Business Machine Corporation under the Hollerith Patents were machines which would perform the operations indicated by the multiple regression equations when such equations were adapted in certain ways.

We will describe the use of these machines by describing their use in a particular instance.

We wished to make success predictions in the junior college field in each of the following subject groups: Physical science and mathematics, English, social studies, languages, and biological science, on the basis of the number of A's and B's earned in high school, the scores in the Thurstone Psychological Examination, Edition of 1929, and the scores on the various parts of the Iowa Content Examination, all recorded at the beginning of the junior college course.

The correlation of any one of these items with college success is usually too low to be used with much efficiency in prediction (exceptions to this general statement are shown by our own data for physics

[1] Hull, C. L.: "Aptitude Testing," p. 487.

and mathematics), and therefore a combination of scores or ratings into a composite rating is desirable. With a variety of correlations of the separate items with success in a college subject present, it is desirable to weight the items differently from the natural weighting. The method which gives this best weighting is the multiple regression equation technique.

For the purpose of this description we shall limit ourselves to the prediction of the first two of our college subject groups. The criterion of success in the subject groups is the mark made in the subject groups the first semester in junior college. The correlations between the items and the two college subject groups are given in Table I. The intercorrelations between the tests are given in Table II.

TABLE I.—CORRELATIONS BETWEEN PREDICTION ITEMS AND CRITERIA

| Criterion | Entrance units (A's & B's) | Thurstone | Iowa | | | |
|---|---|---|---|---|---|---|
| | | | English | Mathematics | Science | Social studies |
| Mathematics and physical science | .136 | .699 | .720 | .537 | .573 | .400 |
| English | .551 | .401 | .356 | .018 | .090 | .127 |

TABLE II.—INTERCORRELATIONS OF ITEMS

| | Thurstone | Iowa | | | |
|---|---|---|---|---|---|
| | | English | Mathematics | Science | Social studies |
| Entrance units | .261 | .320 | .064 | .068 | .242 |
| Thurstone | .... | .560 | .510 | .441 | .468 |
| Iowa English | .... | .... | .223 | .377 | .738 |
| Iowa mathematics | .... | .... | .... | .604 | .221 |
| Iowa science | .... | .... | .... | .... | .457 |

Upon the basis of these correlations the $\beta$'s for the multiple regression equations were obtained.

In order to use the Hollerith card which is used in these machines it was necessary to reduce all scores to a basis of 0 to 9 inclusive, corresponding to the numbers on the Hollerith card. This was done by reducing all distribution of scores to a basis of 0 to 9 according to the schedule of Table III.

TABLE III.—SCHEDULE OF TRANSMUTED SCORES

| Sigma | Transmuted score | Sigma | Transmuted score |
|---|---|---|---|
| $-\infty$ to $-2.0$ | 0 | 0 to .5 | 5 |
| $-2.0$ to $-1.5$ | 1 | .5 to 1.0 | 6 |
| $-1.5$ to $-1.0$ | 2 | 1.0 to 1.5 | 7 |
| $-1.0$ to $-$ .5 | 3 | 1.5 to 2.0 | 8 |
| $-$ .5 to 0 | 4 | 2.0 to $\infty$ | 9 |

The multiple regression equation which is in general[1]

$$\overline{X}_0 = \frac{\sigma_0}{\sigma_1}\beta_{01.23\ldots n}X_1 + \frac{\sigma_0}{\sigma_2}\beta_{02.134\ldots n}X_2 \ldots + \frac{\sigma_0}{\sigma_n}\beta_{0n.1234\ldots(n-1)}\frac{\sigma_0}{\sigma_n}X_n + C$$

where

$$C = M_0 - \beta_{01.23\ldots n}\frac{\sigma_0}{\sigma_1}X_1 - \beta_{02.13\ldots n}\frac{\sigma_0}{\sigma_2}X_2 - \beta_{0n.123\ldots(n-1)}\frac{\sigma_0}{\sigma_n}X_n$$

will be simplified because all sigmas are equal since all distributions have been reduced to a sigma basis. The equation becomes

$$\overline{X}_0 = \beta_{01.23\ldots n}X_1 + \beta_{02.134\ldots n}X_2 \ldots + \beta_{0n.123\ldots(n-1)}X_n$$

$C$ disappears because

$$M_0 - \beta_{01.23\ldots n}X_1 - \beta_{02.134\ldots n}X_2 \ldots - \beta_{0n.123\ldots(n-1)}X_n = 0$$

The $\beta$'s resulting will be in the form of a decimal fraction. We have arbitrarily changed these fractions to whole numbers ranging from 0 to 3 inclusive in order to fit out tabulating equipment. The equation for physical science and mathematics is as follows:

$$X_0 = 0X_1 + 3X_2 + 3X_3 + X_4 + X_5 + 0X_6$$

where the symbols have the following meaning in transmuted score.

$X_1$ = Number of A's + B's made in high school.
$X_2$ = Thurstone test score.
$X_3$ = English test score of Iowa High School Content Examination.
$X_4$ = Mathematics test score of the Iowa High School Content Examination.
$X_5$ = Science test score of the Iowa High School Content Examination.
$X_6$ = Social studies test score of the Iowa High School Content Examination.

[1] Kelley, T. L.: "Statistical Method," p. 283, formula 243.

The equation is reduced to

$$\overline{X}_0 = 3X_2 + 3X_3 + X_4 + X_5$$

since $0X_1$ and $0X_6$ drop out.

Because of changing fractions to whole numbers and restricting the weighting to 0 to 3 inclusive, the multiple correlation coefficient is reduced somewhat. The multiple correlation coefficient of the above equation for mathematics and science is .80. Before the reduction and restriction above mentioned, it was .84.

The equation for the prediction of English was

$$\overline{X}_0 = 3X_1 + 2X_2 + X_3.$$

The multiple correlation coefficient for this equation is .62. Before the reduction and restriction was made, it was .65.

The method of using these equations with the tabulator and key punch is as follows:

A Hollerith card is punched for each student. The card used is one divided into sections of three columns each. In each section the transmuted score of the test is punched in every column. A sample of the card is given here for a student who made a score of four in high school marks, a score of five in the Thurstone test, a score of six in the English part of the Iowa test, a score of five in the mathematics part of the Iowa test, a score of four in the science part of the Iowa test and a score of seven in the social studies part of the Iowa test.

Cards so punched are put into the tabulator and the wiring adjusted to fit the equation. For physical science and mathematics the wires will be connected as follows: Three to the Thurstone score section ($3X_2$), three to the English part of the Iowa test ($3X_3$), and one each to the mathematics part of the Iowa test ($X_4$) and the science part of the Iowa test ($X_5$). With this wiring the cards can be run through the tabulator and the weighted scores for each student will be added up and the sum printed together with the number of the student. In case the tabulator is not a printer or one which adds up the different fields, the numbers can be added up mentally and the result together with the student's number written down. The composite score so obtained may be used directly or transmuted into the scholarship scale used in the institution. Where the multiple correlation coefficient is not very high the regressive effect of the equation should be accounted for in any transmutation, that is, the original spread of the scores must be restricted to fit the effect of the regression equation.

| Junior college success prediction | Pupil number | High school marks | Thurstone test | Iowa high school content examination | | | | Oper. |
|---|---|---|---|---|---|---|---|---|
| | | | | Part I | Part II | Part III | Part IV | Long Beach City Schools |
| 0 0 0

In case the multiple correlation coefficient is high the regressive effect being small can be disregarded. In this way the cards can be run through the tabulator again and again until a prediction is obtained for each subject.

The method described should be of value not only where large numbers of predictions are desirable for use in individual guidance but also for purposes of homogeneous classification upon the basis of several different items. With the data punched on the cards at convenient times during the semester, the cards can be run through and a new homogeneous classification determined for various subjects within a comparatively short length of time.

# A TECHNIQUE FOR EXPERIMENTATION ON GUESSING IN OBJECTIVE TESTS

LOUIS GRANICH

College of the City of New York

At a time when the pedagogical status of the true-false test still "hangs in the balance,"[1] there would seem to be a considerable field for a technique peculiarly adapted to experimentation on the element of guessing. The writer's "index of guessing" is obtained by scattering, among the genuine questions of a true-false or multiple choice examination, a number of questions which involve very obscure facts, or newly coined names. As applied here, to a group of college juniors and seniors, it gave considerable evidence of its validity; it was successfully employed to reduce guessing, proving more effective than the usual instructions against guessing; and it was suggestive of a new analysis of responses to true-false questions.

## Procedure

A fifty item true-false test on general psychology, containing at irregular intervals ten index questions, was administered to a lecture class in social psychology. Attached to each sheet was a pre-test of twenty items of the recall, or completion type, based as nearly as possible on the same topics as the true-false test.

The instructions preceding the true-false test differed for half of the papers. The control group (Group A) received the following full instructions against guessing:

### Read These Instructions

Put a circle around Right or Wrong. You are not expected to recognize all of these statements. *Do not guess*, because of the usual scoring system.

The remaining half of the class (Group B) found these new directions on their sheets:

---

[1] Ruch, G. M.: "The Objective or New-type Examination." Scott, Foresman and Company, 1929, p. 365. See pp. 318–368 for an excellent digest of experimentation on the subject discussed here. Note that probably the best practice in true-false testing involves instructions against guessing, and scoring which corrects for guessing errors.

READ THESE INSTRUCTIONS

Put a circle around Right or Wrong. You are not expected to recognize all of these statements. *Do not guess,* because of the usual scoring system.

Suppose you met this question: R W The Funk Charts measure the discrimination of emotions. There are no Funk Charts, and if there were you could not have heard of them. To answer this question would indicate guessing on your part. There are a good number of questions of this type and of similar types scattered through this test. Each one will indicate guessing if answered.

Therefore answer only what you know.

To obtain a random halving, the sets were arranged alternately according to the two types of directions and were dealt out thus. Each student received a complete test set as he entered. All directions supposed necessary were printed on the sheets, and no questions were permitted across the room, for obvious reasons.

The groups were, then, random halves of a highly selected group, each half numbering thirty-four. They were expected to differ on the true-false test because of the new instructions. The pre-test was considered advisable, in order to confirm the similarity of the two groups as to distribution and total amount of information.

The method used here for establishing a control was preferred as one which resembled the procedure of an ordinary classroom test, and did not smack of experimentation to the students. Every effort was made to give the test under classroom conditions. Several other considerations pertaining to later interpretations also caused the test to be given in the form described.

The scores of the two groups on the pre-test ranged from six to twenty. With a maximum possible score of twenty, the means were 12.56 with a PE of 2.1 for B and 12.23 with a PE of 1.9 for A. The groups appear similar enough to render significant any considerable differences which might be found on the true-false tests. It should be noted that for most of the mean amounts given below the deviations were large; and where differences occurred, overlapping between both groups was considerable.

Time was not called until it appeared that practically every student had finished. The students were then asked to indicate the questions on which they had guessed. Precautions were taken to secure complete cooperation. The students were addressed informally by the writer, who is not their instructor,[1] and were told that the

[1] I am indebted to Dr. Maximilian R. Schneck of the College of the City of New York for permission to administer this examination.

test in no way concerned themselves, that its purpose was purely experimental. They were asked to tear off or delete their names, or any other identifications, from their papers. The point was repeated to them that in spite of their directions they certainly had guessed on some questions, probably on a large number. They were asked to cooperate in an "interesting experiment" and to indicate by a capital *G* all guessed answers. Papers were then collected.

The writer later had the opportunity to receive objections from the class against the procedure of the experiment. A number of students questioned the ability of the class to indicate its own guessing. No opinion was met that lack of comprehension, suspicion, or resentment had affected the frankness of the class.

## Summary of Results

Evidence of the cooperation of the class is found in Table I. It will be seen that in Group A, seventy-two per cent of those questions called "guessed" were actually correct, while in Group B, seventy-five per cent were correct. It may be safely assumed that the class tended to be unreserved in indicating guessing; which was the effect desired. And yet both groups indicated as guessed only half of the index questions which they answered. The explanation of this apparent paradox

Table I.—Mean Scores of Group A (Control) and Group B (New Directions)[1]

| | Group A (34) | Group B (34) |
|---|---|---|
| Pre-test scores | 12.23 (PE 1.9) | 12.56 (2.1) |
| Index questions answered | 4.2 (PE 1.7) | 1.9 (1.2) |
| Index questions indicated as guessed | 2.1 | 1.0 |
| Actual questions right | 27.2 | 25.1 |
| Actual questions wrong | 6.3 | 5.1 |
| Actual questions omitted | 6.5 | 9.8 |
| Actual questions indicated as guessed and correct | 3.8 | 3.3 |
| Actual questions indicated as guessed and incorrect | 1.5 | 1.1 |

[1] Note the size of the PE's of the index distributions for both groups. The difference between the pre-test means bears a ratio to its P.E. of about .75, making it of significance. As for the index means, the ratio of the difference to its P.E. is about 6.5, easily demonstrating statistical significance.

will appear later. It is evident, however, that the student's introspection makes a poor index of guessing.

With reference to the decrease in guessing, it will be seen from Table I that omissions were increased forty-nine per cent. The number wrong was decreased nineteen per cent, the number right eight per cent and the number indicated as guessed seventeen per cent. Comparison of this last figure with the others shows that Group B was readier to admit guessing than was Group A. It should be stressed that all results here obtained are based on a single test and on a highly selected group; so that the reliability of these differences is not finally established for all groups. To judge by the mean number of index questions answered, guessing was reduced about one-half. However, it may be argued that since Group B was put on guard against these questions, the results on the index may not be valid as a measure of guessing—that the reduction was caused by the recognition of the index questions as such, and did not indicate a reduction in guessing on the body of the test. Three evidences are offered in answer:

1. Argumentative evidence which will be presented below in the analysis of student responses.

2. Each group marked as guessed almost a precise half of the index questions answered. As a matter of fact, it will appear from Table I that Group B, as compared with Group A, tended to overestimate the number of its guessed answers. This has been pointed out above. Group B was therefore less aware of guessing on those index questions which it answered than was Group A; since if they were equally aware of guessing, and readier to indicate it after the test, they should have indicated a greater proportion as guessed than did Group A. When Group B, therefore, guessed on an index question, it was at least in part due to the fact that they were relatively unaware of guessing. The relative tendency against guessing carried over from the body of the test, and explains wholly or in part the reduction in the average index score.

3. Fairly conclusive evidence is furnished by these figures: For Groups A and B, the ten highest and the ten lowest pre-test scores were selected, and their index scores compared, as shown in Table II. The pre-test scores were used as the criterion of achievement, since the true-false scores were subject to varying factors and were therefore unsatisfactory. It will be seen that the best and the poorest students scored about the same on the index. For Groups A and B the ten papers were then selected which had the highest and lowest

number of index questions answered, and their pre-test scores were compared, as shown in Table III. It will be seen that the students who guessed most and least on the index were equal in ability. These figures indicate roughly that no significant relationship exists between knowledge of the material which has been studied and ability to discern the index questions. An additional conclusion which is bound up with this latter would have to be, that the better students do not differ from the poorer in their tendency to guess, or rather to attempt questions which are new for them. Such questions of course constitute less of sheer guessing for the superior pupils. Furthermore, the absolute number of such questions is greater for the poorer pupils.

TABLE II.—MEAN SCORES OF FORTY STUDENTS SELECTED ACCORDING TO ACHIEVEMENT

| | Group A | | Group B | |
|---|---|---|---|---|
| | Highest ten | Lowest ten | Highest ten | Lowest ten |
| Pre-test | 15.5 | 9.2 | 16 | 8.3 |
| Guessed (index) | 4.5 | 4.5 | 1.6 | 1.7 |
| Guessed and indicated as guessed | 2.6 | 2.3 | 1.1 | .9 |

TABLE III.—MEAN SCORES OF FORTY STUDENTS SELECTED ACCORDING TO INDEX SCORES

| | Group A | | Group B | |
|---|---|---|---|---|
| | Highest ten | Lowest ten | Highest ten | Lowest ten |
| Index score | 7.5 | 1.7 | 4.1 | .3 |
| Index questions indicated as guessed | 3.5 | 1.2 | 2.4 | .3 |
| Pretest | 12.4 | 12.0 | 12.4 | 12.4 |

The validity of the index, then, is apparently not affected by awareness on the part of the student or by his knowledge of the subject covered. The index does not however purport to be an absolute measure of guessing, since it contains a certain amount of

recognizable material.[1] Different index scores will result from different sets of index questions. But if the items are equated as to their indicative values, they form a basis for precise comparisons, and for correlative work. They are obviously as applicable to the multiple-choice test as to the true-false.

## The Index Questions

The ten index questions are given here in the order in which they appeared on the test, followed by figures representing the responses elicited by them of right or wrong, and also the number of times they were indicated as guessed.

It will be seen that the guessing element in these questions varies from an almost absurd amount in several, to so small an amount in others as to render them almost conventional. A better index would be composed of equated items. It is not sheer guessing that is measured by most of those here, but the tendency to attack material that is new, and offers no legitimate basis for an answer.

---

[1] Nor does it measure those questions which are answered with an even chance of success. If it did, the following formula could be applied:

$$\frac{N}{I-N} \times O = G$$

where $G$ is the number of questions guessed, $I$ the number of questions composing the index, $N$ the number of these questions answered, and $O$ the number of omissions on the body of the test. Half of $G$ would represent the number right or the number wrong through guessing. If the formula be applied to our data above, a striking similarity is found between both groups, in their estimated "true" scores. The number right or wrong through guessing is 2.4 for Group A and 1.2 for Group B. Subtracting these amounts from the number right to find the "true" score, we obtain 24.8 for A and 23.9 for B. Subtracting from the number wrong to find the actual amount wrong for other reasons than guessing, we obtain for both halves an identical remainder of 3.9. This is what we should expect for such similar groups. Furthermore, the ratio of their $G$'s is about equal to the ratio of their index scores. The explanation of these agreements, however, presents many difficulties. They are probably spurious, unless by accident the index sampled guessing correctly. The possibility of using the index as a refinement in scoring, even for groups as a whole, requires further demonstration.

It may be argued that the index questions, if equated, contain a constant proportion of guessing, and can be converted to pure guessing scores by multiplication. As will appear later, however, what we encounter on the test is not pure guessing, but guessing dispersed as an element among many questions, and subject to uncertain rules of probability.

The questions are composed either of obscure facts or of propositions involving a newly coined term, and should be conventional in structure. The practical application of the index and of the new instructions, remains to be further demonstrated. It may be that the present standard instructions (*Do not guess unless you have some basis for an answer*) will offer the best results, and that guessing need not be further reduced. It may be that R-W scoring is more valid for the very reason that it takes into account the wrong answers. It will prove too difficult to construct tests with reliable individual indices, and too cumbersome to score such tests, to allow of their widespread use. The use of meaningless material may prove objectionable in classroom practice since it renders possible retention of matter which is undesirable. It may also be said that the new instructions are threatening and punitive in attitude. But they are no more so essentially than instructions which warn of R-W scoring; in fact they offer more reasonable justification to the student for penalizing. Whatever pedagogical faults may appear against the technique, there remain many possibilities for its use in objective tests outside of the

TABLE IV.—SHOWING THE EFFECTIVENESS OF THE VARIOUS INDEX QUESTIONS

| | Marked right | Marked wrong | Indicated as guessed |
|---|---|---|---|
| 1. Convergence of factors will probably explain the transfer of training | 14 | 12 | 9 |
| 2. The Juke's history shows that fourteen per cent did not become public charges | 22 | 10 | 14 |
| 3. The acquisitive instinct is today considered conductive as well as dynamic | 10 | 8 | 8 |
| 4. One cause of illusions is the functioning of diffused association-arcs | 23 | 1 | 12 |
| 5. There are never more than three thalamic membranes | 4 | 4 | 7 |
| 6. Glandular misfunctioning is the cause of pitrinosis | 9 | 0 | 6 |
| 7. Dividing a learning curve into zones increases reliability but lowers validity | 10 | 10 | 11 |
| 8. There are more sensitized pigment spots in the brain than in the spinal cord | 17 | 7 | 10 |
| 9. Laselles found the white and colored races about equal in keenness of perception | 17 | 11 | 18 |
| 10. Jung was the first psycho-analyst to refute the existence of the negative personality | 9 | 9 | 9 |

classroom. It will be noted that the possibility of detection constitutes adequate motivation for reducing guessing.

The distribution of index scores (Table V) shows that extreme amounts of guessing are almost entirely eliminated, while there is a much greater frequency of very small amounts.

The writer found these questions easier to prepare than those which were real. Care must be taken only that obscure facts be obscure enough, and that an answer of *False* should not by some chance be an honest and correct answer to a sentence involving a neologism, as for example, in "The acephelon is located in front of the thalamus."

The greatest utility of the technique lies in further theoretical investigation. Among the possibilities are: (1) Correlative work and the determination of relations between guessing and the different elements of a test situation. (2) Comparison of the amount of guessing going into different types of tests, including even completion tests; also, estimation of the relative amount of guessing going into specific tests. (3) A study of the distribution of student-types as regards guessing. (4) Analysis of characteristics of questions which invite guessing. (5) Study of the extent to which students are aware of their guessing. (6) Retention of content with different tests. (7) A measure of one phase of classroom discipline, and of the effectiveness of different instructions, motivations, personalities, etc. (8) Experimentation leading to the perfection of the index, confirming its validity, and determining its possible application to the classroom. The possibility of employing it as a refinement in scoring is of greatest importance here.

The dilemma appearing from our results is this: While only half of those index questions answered are indicated as guessed, three-quarters of those so indicated on the test proper are correct. It is of course assumed that the effect of "specific determiners" alone will not explain this. The analysis which will presently be given is suggested in explanation of this phenomenon. It does not, of course, derive any definite corroboration from our figures, but is merely suggested by them.

TABLE V.—DISTRIBUTION OF FREQUENCIES OF GUESSING

| Index score | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Frequency, Group A..... | 1 | 3 | 5 | 5 | 10 | 2 | 0 | 3 | 2 | 2 | 1 |
| Frequency, Group B..... | 8 | 10 | 6 | 6 | 1 | 2 | 0 | 0 | 0 | 1 | 0 |

## The Nature of Responses

Discussion to date has assumed that the response to any question is determined by one of several factors; either information, guessing, misinformation, or a "hunch" resulting from the characteristics of the question (Wiedemann's *specific determiners*). It has been recognized that guessing varies at different times, with different groups, under different test conditions, and also with each individual examinee. It has been found that guessed answers are more often correct than incorrect;[1] and this fact has been explained on the basis of subliminal configurations of knowledge. There is also evidence that R-W scoring over-penalizes students;[2] and it has been pointed out, in partial explanation of this fact, that answers wrong through misinformation are penalized doubly. It is obvious that this routine analysis will not suffice to explain our results in this study. A closer analysis of the factors determining a response is therefore relevant.

Questions are *factual* insofar as they measure the recognition of terms or of facts. Questions are *deliberative*, involving *judgment*, to the extent that they require the recall of facts to decide the truth of a statement not already taught. A question intended as factual may sometimes be attacked in deliberative fashion, to substitute for or to supplement recognition. "Deliberation" involves recall, comparison, criticism, ingenuity, application, practice in problem attack, and any number of phases of general intelligence. It may be an acceptable compensation for inability to recognize, if by its means the correct answer is arrived at, from true premises and by a correct process of reasoning.

The answer to any question is the result of a number of factors, not of one. Thus in the (rather poor) question, *Wundt established the first psychological laboratory in 1879*, a student may recognize the name *Wundt* (information) but be uncertain as to whether this man was the pioneer in question (guess). He has an idea that Titchener was the founder (misinformation). He does not recognize 1879, but decides its probable accuracy by checking it with other dates (deliberation, as well as guess). The probable accuracy of the date, and other considerations, now lead him to believe the entire statement true (hunch, or suggestibility). He answers in spite of his uncertainty (mental set as to guessing from directions).

[1] West, P. V.: A Critical Study of the Right Minus Wrong Method. *Journal of Educational Research*, Vol. VIII, 1929, pp. 1–9.

[2] Ruch, G. M.: *Op. cit.*, p. 352.

We should then regard guessing as a *factor* insofar as a question is answered with uncertainty; while misinformation becomes an inadequate term. There may be misinformation, it is true, as to a part of a question; problematic or partly problematic questions answered with certainty but incorrectly are fairly frequent. But practically all instances of misinformation, as defined hitherto, are accompanied more or less by uncertainty. It is difficult to assume that students learn complex facts in a form precisely incorrect.

If the final score formula for the number right can not be expressed in terms of whole questions, does it tend to be the sum total of the *fractions* of information plus one-half the total of *fractions* of guessing-factors, plus approximately one-half of the *fractions* of suggestibility factors? Or does the final score depend on the proportions in which these factors are combined? Does each sentence bear its own rules of probability? (Note that the recognition of a single term in a statement may furnish a clue to the correct answer; for example, by elimination.) Do intelligence, training in problems, and application enter also as factors compensating, legitimately or not, for lack of knowledge? Is much recognizable material necessary to assure a student that he is not guessing? Above all, what rules of probability govern questions made up of different recognizable, unrecognizable, and problematic proportions?[1] Would these rules explain the over-penalization of R-W scoring? Further study might well be directed along these lines.

It need not be pointed out that many questions, especially shorter ones, are entirely recognizable or unrecognizable. It is also true that there are degrees of learning; that some facts are on the borderline of memory. This situation introduces other factors which affect scores: application, amount of time, and accident in striking upon effective associations. These last few items will be recognized as universal causes of unreliability, together with such elements as copying, ambiguity of the questions, attitude of student, conditions of the test, obscure wording, etc. In problematic questions an analogue to guessing is found in the process of hitting upon a line of reasoning which leads to a correct answer, although the background of information is very limited, and the judgment poor.

---

[1] Wood found relationships between the problematic quality of questions and their reliability, as well as one aspect of their validity.

Wood, Ben D.: Studies of Achievement Tests. *Journal of Educational Psychology*, Vol. XVIII, 1926, pp. 10 and 13.

At any rate, we cannot regard some one factor as the single cause of a given response, nor can we consider such responses at all analogous to coin-flipping. An exception may perhaps be made for simple, short, factual questions.

We may now be able to explain the peculiarities of true-false tests which have already been mentioned. We know, for example, that enough uncertainty may enter into a response for the student to call it guessing. And yet such questions may contain enough recognizable material to be answered correctly in more than half the cases. We may now reject the dubious concept of misinformation accompanied by certainty. We may also explain why answers which are changed by the student are correctly changed most of the time.[1] We may deduce from our theory such very plausible propositions as these: A multiple choice of four answers may really offer only two actual alternatives, since two others may be conspicuously absurd. Or, if one of the answers is known to be incorrect, the chance of guessing correctly will be one out of three. If, on the other hand, none of the alternatives are recognized, but one of them offers a "catch," then the chances of correctness are less than one out of four.

It is obvious that if a student does not recognize a fact, he can not remember whether that fact was ever discussed in class, and can not decide whether he is expected to know it. Unless the question is altogether absurd or unconventional in appearance, he can not be sure whether it is an index item, or one which he should know. It is for this reason that knowledge of the presence of the index does not affect its validity. This analytical evidence the writer has already promised to present. The effectiveness of the new directions is due to the same situation. To guess on an item unrecognized in part is risking an increase in index score.

## Summary

1. The index offers a valid measure, although not necessarily absolute, of guessing on objective tests. Guessing is here defined as the tendency to answer questions which are unrecognized either wholly or in part, when an answer can not be deduced with certainty from such information as the student possesses.

---

[1] Mathews, C. O.: Erroneous First Impressions on Objective Tests. Also, Lowe, M. L. and Crawford, C. C.: First Impression vs. Second Thought in True-false Tests. *Journal of Educational Psychology*, Vol. XX, 1929, p. 195 and p. 286 respectively.

2. Warning as to the presence of index questions in the test used here proved to be more effective in reducing guessing than the usual full instructions and warning of penalization.

3. For a number of reasons already explained, students can not reliably indicate their own guessed answers. This fact is demonstrated by comparison of index scores with our other numerical results.

4. In this study, superior students did not differ from poorer students in their tendency to guess; that is, in their average index scores. It follows that they were no better able to discern index items even when aware of their presence. The students who guessed most and those who guessed least, according to their index scores, were nevertheless equal in achievement according to a pre-test.

5. Some of the results of this study were difficult of explanation in terms of the conventional analysis of student responses to true-false questions. A revised analysis is presented, in terms of which a plausible explanation can be found for many of the peculiarities of true-false tests, besides those appearing in this study.

6. Argumentative evidence, based on this analysis, is produced to show that responses are more often the result of several psychological factors than of one. Simple factual questions are an exception to this rule.

7. Furthermore, such answers are not subject to simple mathematical rules of probability in scoring; their probabilities of correctness vary according to the different proportions of guessing, information, etc., that enter into the responses.

8. Further studies employing the index technique are suggested above; and studies leading to the criticism and improvement of the index will be of value.

# NEW PUBLICATIONS IN EDUCATIONAL PSYCHOLOGY AND RELATED FIELDS OF EDUCATION

CONDUCTED BY FRANCES M. FOSTER

*The Science of Psychology*, An Introductory Study, by Raymond Holder Wheeler. New York: Thomas Y. Crowell Co., 1929. Pp. XI + 556.

In "The Science of Psychology" Professor Wheeler of the University of Kansas makes a pioneering attempt to present general psychology in textbook form from the Gestalt point of view. In spite of the difficulty of the task, the result is neither one-sided nor incoherent. Nor is it (perhaps more surprisingly) unduly abstruse or difficult. In parts the discussion involves certain inherent difficulties, but the style is simple and the presentation logical and clear.

Whereas most textbook writers seem to base their eclecticism upon the principle that those explanations which are most widely accepted at the time of writing are to be preferred, Wheeler seems to base his upon the principle that those explanations which appear to him simplest and most inclusive, are to be preferred. It would be more accurate to say that in both Wheeler's book and in the usual textbook there is a mixture of these two principles, but that in the latter there is more evidence of the principle of general acceptance than of the principle of "parsimony," whereas in Wheeler's book the reverse is the case.

All living activity is described by Wheeler as goal activity, governed by the "Law of Configurations" and the "Law of Least Action." A goal he describes as

> the termination or end of a given response, however simple or complex . . . Represented by those stimuli or objects in a stimulus-pattern, the reaching of which relieves tension in the organism. Represented in the organism by a relative equilibrium in the neuromuscular system . . . In bringing behavior into relation with dynamical principles, the goal is interpreted as a point in an energy system toward which a body moves under stress; it is the point of low stress of the system, and the organism is regarded as part of the system in which it moves . . . The situation is akin to a gravitation system with its center constantly moving, along with shifts in the alignment of potentials elsewhere in the system.

Wheeler's description of the "Goal" concept is more definite and concrete than McDougall's description of the concept of purpose and less restricted in scope than the concept of desires proposed by the psychoanalysts. Although he employs it as an explanatory principle quite as universally as these other psychologists, he does not give the impression of distorting the facts to prove his point of view. On the contrary, the large amount of experimental data which he brings together in the 556 pages of his book is clarified and unified by consistent examination from his defined viewpoint.

Wheeler progresses systematically from the sociological aspects of psychology to the biological aspects, with the chief emphasis, where it belongs, midway between these two extremes, on "Emotive Behavior" and "Learning Behavior," the treatment of the latter being so particularly thorough and stimulating as to make the book an excellent reference or text in advanced courses in educational psychology.

His treatment, in the first chapters of the book, of "Social Behavior and Its Conditions" is without distinction—just another compilation, not unusually coherent, of the facts and theories generally reported in social psychologies. It has the merit of being thoroughly modern and of containing some valuable and original formulations of principles. The latter give the impression, however, of being merely inserted for good measure without affecting the choice of other material in this section of the book.

But this deficiency in the first part of the book is in sharp contrast with the rest of the book, and since the first part is fairly typical of that vast number of texts whose chief reason for existence is their clarity and modernity, it forms a handy reference point by which to judge the superiority of the body of the work.

JOHN N. WASHBURNE.
Syracuse University.

---

*Manual for Determining the Equivalence of Mental Ages Obtained from Group Intelligence Tests*, by Ross O. Runnels. Test Method Helps, No. II. Yonkers-on-Hudson, N. Y.: World Book Co., 1930. Pp. 14, paper.

*Expressing Educational Measures as Percentile Ranks*, by F. C. and O. K. Buros. Test Method Helps, No. III. Yonkers-on-Hudson, N. Y.: World Book Co., 1930. Pp. 27, paper.

It has long been noted by users of tests that there are gross differences in mental ages and intelligence quotients obtained from

different group tests. Due to variations in methods of standardization and differences between the groups by which the various tests were standardized, a sad lack of equivalence exists between the tables of norms. Dr. Runnels administered the National, Haggerty, Otis-advanced, Terman Group and Otis Self-administering Tests to 1422 pupils in Grades V to IX inclusive. The equivalence of scores was determined by equating percentile ranks. For each percentile rank, the average of the mental ages given by the manuals for the various tests was taken as the new mental age norm. Of course, no such procedure will increase the reliability of a group test, but Runnels' tables will be of great assistance in obtaining more comparable results when a variety of group tests is used in a school in successive years.

"Test Method Helps, No. II" gives an admirably clear and simple explanation of ranks and percentile ranks and includes tables to aid the computation of percentiles when the number of cases is from 11 to 100. For those who are not adept in the use of a slide rule, this little book provides a quicker and easier method of computing percentile ranks than has hitherto been available.

LAURANCE F. SHAFFER.
Carnegie Institute of Technology, Pittsburgh.

---

*Child Adjustment*, by Annie Dolman Inskeep. New York, D. Appleton and Co., 1930. Pp. XVI + 247.

A study of the facts and principles presented in "Child Adjustment" will enable the teacher and educational specialist to gain new insight into the educational process as related to the growth process in the school child. The book serves to bring together the major findings of many scattered bits of research. As the author states in the preface, the essential purpose of the book is that of "finding out how the child's body, mind, and emotions differ from an adult's, how they develop into the adult stage, and how they should be cared for during school years." The topics discussed include adjustment with reference to height and weight, growth and functions of the larger muscles, handedness and footedness, the internal organs, different developmental ages, the teeth, adjustment in relation to growth of brain and nerves, the eyes and ears, adjustment in relation to growth, development of the mind, measurement of intelligence, critical adjustment periods, emotional health and mental hygiene. In the appendix there is a

discussion of different psychological theories, behaviorism, gestalt psychology, Freud, Jung, Adler, psychoanalysis.

The author shows originality in the methods devised for studying different factors in child behavior and makes the material of practical usefulness through her descriptions of actual school practice in studying the physiological, mental and emotional development of pupils. The best sources of scientific data have been consulted and are referred to in each chapter. The interrelation of physical and mental factors in the child's development is emphasized with each topic discussed. The belief is expressed that if the child's home and school adjustments have been made with his physical, mental and emotional needs in view, the child makes better progress in his school work. Cases of discipline, failure, nervous break-down, may be attributed to failure to understand the child as different from the adult.

In occasional instances the reader may feel that the author lays too much stress on physical deficiencies as major causes of maladjustment, as for example when lying and stealing are found in conjunction with adenoids or poor eyesight, and the behavior defects are attributed to the physical ones. A tendency to attribute school failure to physical defects is implied in the author's discussion, but there is also emphasis throughout the book on the importance of investigating all causal factors in an adjustment problem.

GERTRUDE HILDRETH.

The Lincoln School of Teachers College.